Are Active Norwegian Mutual Fund Managers Paid for Luck or Skill?

A New Approach to Distinguish Between Luck and Skill Among Active Norwegian Mutual Fund Managers

KRISTIAN FJÆRBU LØHAUGEN VIKSTØL

SUPERVISOR
Valeriy Ivanovich Zakamulin

University of Agder, 2021
School of Business and Law
Department of Economics and Finance

Master

Abstract

This thesis investigates whether the performance of actively managed Norwegian mutual funds is due to luck or skill. First, we measure the performance on aggregate and individual levels using the Carhart (1997) four-factor model. We then turn to the bootstrapping procedures, free from data snooping bias, proposed by Kosowski, Timmermann, White, and Wermers (2006), Fama and French (2010), and a new approach proposed by Harvey and Liu (2020) in evaluation of luck or skill among the Norwegian mutual fund managers. Through the data, free from survivorship bias, aggregate fund performance suggest no statistically significant evidence of abnormal risk-adjusted net-of-fees return. At the individual level we find both positive and negative statistically significant abnormal net-of-fees returns, risk-adjusted. Finally, there is no statistically significant evidence of skill amongst the Norwegian mutual fund managers. We do, however, find evidence of a lack of skill among the bottom performing managers.

Preface

This master thesis concludes the studies in a master's program of finance at the School of Business and Law, UiA. The topic of this thesis was chosen due to my interests in financial markets and specifically the Norwegian mutual fund market. Investments into the financial markets made by private persons have been growing in Norway. From this, an opportunity to investigate whether the actively managed funds can generate returns warranting their fees is a motivation in itself. Also, there is an ongoing debate as to whether one should invest in passive funds or actively managed funds. Attaining more knowledge, and investigating further into the subject is thus of great interest to me.

During this thesis, I have faced both interesting and educational challenges. Including how the Norwegian mutual fund market is organized, statistical methods to measure the performance of funds, and getting an in-depth view of how mutual fund performance measurement has evolved since the 1950s.

Choosing to write this thesis in English as opposed to my native language, Norwegian, stems from the nature of the topic. Past research and expressions within the field and statistical methods are easier to convey in a meaningful way in the English language. Through this, I believe that I am able to communicate the methodology and past research in a more accurate way than what would be the case if terms and methodologies had to be translated.

Finally, I would like to thank my supervisor Valeriy Ivanovich Zakamulin. He has brought forward insightful information and suggestions throughout writing this thesis.

Contents

List of Tables

List of Figures

1 Introduction

This thesis investigates whether Norwegian mutual fund managers are able to generate abnormal returns for their customers, and whether abnormal performance is due to luck or skill. Beating the "market" has long been seen as impossible. Yet, investors buy into actively managed funds as opposed to passively managed index funds. As investors are paying a higher price in actively managed mutual funds, one would expect a higher return to cover the higher price paid.

Actively managed funds buy and sell stocks in their attempt to generate abnormal returns. That is, they seek to buy stocks which they expect will have a price increase, and sell stockholdings that they expect will not contribute to abnormal returns. Passively managed index funds, however, are low-cost funds, that try to mimic the return of a market index. The active mutual funds measure themselves against benchmarks, which they aim to beat. Mutual funds can have different benchmarks but they are often an index, which can be related to a broad index, sector, or industry.

In order for active mutual funds to generate abnormal returns net of fees, the market cannot be fully efficient as proposed by Fama (1970) in his Efficient Market hypothesis. More specifically, if the active mutual fund managers are able to exploit information from political changes, market trends, the economy as a whole, or other factors that could affect the pricing of assets then the market is not fully efficient. From this, Fama argued that any abnormal returns would solely be attributed to luck, as opposed to the stock-picking skill of a mutual fund manager.

We thus have two questions to answer in this thesis: 1) Are Norwegian mutual fund managers capable of generating abnormal returns net of fees, and 2) Are the Norwegian mutual fund managers skillful enough to warrant their fees, or are they lucky?

To address these questions, and separating this thesis from the before-mentioned studies on the Norwegian mutual fund market. A longer time period is selected for fund returns, in addition to applying the new proposed bootstrapping approach of Harvey and Liu (2020). The sample consists of 107 actively managed Norwegian mutual funds. Further, funds are required to have a minimum of 12 monthly returns in the period extending between January 1987 and December 2019.

The first question of this thesis, more specifically, whether active Norwegian managers are able to generate abnormal returns that are both risk-adjusted and net of fees. Calls for the application of regression estimation using factor models. These models will be discussed in later sections.

From a statistical point of view, we know that a large sample contains both positive and negative abnormal performances, which may be solely due to luck. This is a critical problem in multiple testing, where if you run a test enough times, you will find significant evidence. This is a type of data-mining bias and warrants the application of a method in which can separate skill from luck.

As an answer to the before-mentioned statistical problems in multiple testing, we can utilize the methodology of bootstrapping. This will help us in distinguishing luck from skill, answering the second question of this thesis. In order to separate luck from skill, one needs to test a null hypothesis stating that all funds perform similarly to the benchmark. Specifically, that no funds generate abnormal returns. This test relies on the application of bootstrapping, because we have no other method available to conduct it. Further, we generate the distribution of funds alpha under the null of no abnormal returns, or, equivalently, no skill. From this, we infer whether the alpha generated by the funds is due to luck or skill. The approaches applied are based on those of Kosowski et al. (2006), Fama and French (2010), and the approach which has not previously been applied in Norwegian mutual fund performance measurement before, Harvey and Liu (2020).

In addition, the new method proposed by Harvey and Liu (2020), makes use of a double bootstrap approach, as opposed to a single round bootstrap in the frameworks of Kosowski et al. (2006) and Fama and French (2010). This method also takes into consideration the Type I and Type II error rates of the test statistics applied to the original dataset. From this, one can determine which cutoff test statistic to use. The test statistic should satisfy one's requirements in terms of Type I and Type II errors, in addition to the prior belief of the number of outperforming funds. Furthermore, all these papers employ the Carhart (1997) four-factor model, where alpha is the measure of abnormal return.

The U.S. mutual funds are the most frequently investigated mutual funds, for which the majority of the literature and research is based. Studies concerning the Norwegian mutual funds are few, especially those taking on a broad and extensive investigation as in this thesis.

There is, however, a study on the Norwegian mutual fund market that applies the methodology of Kosowski et al. (2006). This study was conducted by Gallefoss, Hansen, Haukaas, and Molnár (2015). From their results, Gallefoss et al. (2015) claim that the performance of top and bottom funds is attributable to stock-picking skill or a lack thereof, respectively. Contradicting evidence to the findings of Gallefoss et al. (2015) in the Norwegian mutual fund market is found in the study conducted by Sørensen (2009). In his paper, Sørensen applied the methodology of Fama and French (2010) and did not find evidence of skilled Norwegian mutual fund managers. Sørensen (2009) did, however, find evidence of unskilled managers in his study, where the inferior returns of some managers could not be attributed to bad luck alone.

Results of the net-return of Norwegian mutual funds indicate that on aggregate, they are not able to generate statistically significant returns different from zero. As net-return of funds are used, this evidence tells us that the managers are not skillful enough to go beyond covering their costs. On the other side of the spectrum, findings tell us that the under-performers indeed have inferior skill. That is, the returns provided by these funds cannot be attributed to the fund managers having bad luck, but their lack of skill actually generates negative risk-adjusted-performance.

Lastly, addressing test power, which tells us whether or not our analysis is viable. Findings from Harvey and Liu (2020) suggest too low test power in the framework of Kosowski et al. (2006) and Fama and French (2010). In addition, measuring the test power of the cutoff t-statistics found through the Harvey and Liu (2020) methodology suggests too low test power also in this approach. From this, we gather that the Norwegian mutual funds do generate abnormal returns net of fees. We are, however, not able to determine whether this is due to skill or luck, due to the power of the tests.

The rest of the thesis is organized as follows. Section 2 comprises a literature review of studies and academic papers. Next, models and methodology are presented in Section 3. The data and its properties are detailed in Section 4, along with criteria and assumptions. Section 5 reviews the empirical results of applying the models and methodology to the data. Finally, conclusions are shown in Section 6.

2 Literature Review

This section comprises a review of past studies on mutual fund performance, the luck and skill in these performances, and bootstrapping, along with academic literature on the matter. The main goal of the section is to introduce and brief readers to the most relevant and important studies and papers related to the subject, and, in turn, establish expectations for the findings of this thesis. To create a foundation for the understanding of performance evaluation, the first part comprises historical literature essential to this topic. In the subsequent part, a review of the literature and implementation relevant to the bootstrapping procedures measuring luck versus skill will be presented.

2.1 Mutual Fund Performance

Markowitz (1952) introduced the topic of portfolio theory. His idea was that diversification could reduce the risk of holding financial assets. The contribution of Markowitz's papers to the financial research field landed him the Nobel memorial prize in economic sciences in 1990, along with William Sharpe and Merton Miller. Sharpe (1964), Lintner (1965), and Mossin (1966) created the Capital Asset Pricing Model (CAPM). The introduction of CAPM serves as a crucial piece in the field of Economic theory. The model is fundamental in the performance measurement and has been developed to evaluate risk-adjusted returns for mutual funds. This was done by Jensen (1968), where he created a single-factor model.

First, the Sharpe ratio was developed by (Sharpe, 1966) to evaluate the risk-adjusted performance of mutual funds, taking into consideration the total risk of the portfolio, as opposed to the Treynor ratio, taking into consideration only the systematic risk of the portfolio. Further, Sharpe applied this ratio to evaluate U.S. mutual fund performances (Sharpe, 1966). From his results, Sharpe found that a fraction of the outperforming funds continued to outperform. But, from viewing findings as a whole, he concluded that investing in an actively managed mutual fund was a poor investment, as findings suggested that mutual fund managers in his sample seemed to use most of its fees to create diversification rather than finding underpriced assets.

Later on, Jensen (1968) as mentioned, based upon the CAPM framework, developed Jensen's alpha. Jensen's alpha is used to estimate the abnormal performance of a mutual fund. In theory, an actively managed mutual fund should create a positive alpha, whereas a passive index should generate an alpha of zero. Jensen applied this tool in order to evaluate the risk-adjusted performance of mutual fund returns. In Jensen's paper from 1968, he evaluated U.S. mutual funds and estimated their alpha net of fees. From his findings, he concluded that on average, the mutual fund managers where not able to generate positive alpha.

Ippolito (1989) investigated U.S. mutual funds, and his findings lead him to state that mutual funds, net of cost, actually outperformed the S&P500 index. However, in the evaluation of mutual fund performance, the benchmark against which funds are measured is of great importance. The choice of the benchmark can affect Jensen's alpha as Lehmann and Modest (1987) presented evidence of in their paper. As a result, Lehmann and Modest (1987) argued that one needs an appropriate benchmark that will accurately represent the common factors determined by security returns. Finally, Elton, Gruber, and Blake (1996) investigated the findings of Ippolito (1989), and they found that the funds in Ippolito's sample contained a high amount of small stocks not included in the S&P500 index. The small stocks had a significantly high return, which in turn had contributed to the outperformance of funds. Elton et al. (1996) argued that Ippolito (1989) had employed an inappropriate benchmark, which in turn had generated the findings of positive alpha. When adjusting the study conducted by Ippolito, the positive alpha findings turned negative.

Another study critiqued due to the benchmark applied is that of Wermers (2000). Wermers (2000) examined U.S. mutual funds and divided mutual fund performance based on net return and their stock holdings. When investigating the average mutual funds return and the return on stock holdings, he found a difference of 2.3 percent in return. This difference was mostly attributed to expenses and transaction costs. The remainder of the difference could be accounted for as an underperformance of non-stock holdings. Later the same year, Moskowitz (2000) critiqued Wermers (2000) on the use of his benchmarks, and in turn the findings in his paper. Moskowitz (2000) argued that the benchmarks which Wermers' had applied consisted of small and risky firms. The firms included in the benchmark had generally performed poorly over the sample period, which led to a skew in the results. From the application of the benchmark, the findings of Wermers (2000) could have been inflated.

Underperformance of U.S. mutual funds is suggested in the paper of Malkiel (1995). In the period 1971-1991, the returns of the investigated mutual funds did not provide evidence of the fund's being able to "beat the market". The fact that benchmark choices affect findings led to the creation of multi-factor models. The multi-factor models account for market anomalies. Fama and French (1993) introduced one of the most used and most famous multi-factor models, known as the three-factor model. The three-factor model extends the single-factor model created by Jensen (1968) by two additional risk factors to the market factor, size (SMB) and value (HML). Later, Carhart (1997) added a momentum factor to the three-factor model of Fama and French, creating the Carhart four-factor model. The one-year momentum factor that Carhart (1997) added was originally developed by Jegadeesh and Titman (1993).

Gruber (1996) evaluated mutual funds between 1985 and 1994. He applied a four-factor model, consisting of excess market return, the difference in return between small- and large-cap portfolios, a factor measuring the difference in return between high growth- and a growth portfolio, and finally, the excess return of a bond index. The findings of Gruber (1996) suggested an underperformance of mutual funds when compared to a weighted average of indices. In the same paper, however, Gruber (1996) investigated the funds gross of expenses. From this, he argued that the fund managers had abilities to generate abnormal returns, and as such, the managers had stock-picking skills. The stock-picking skill did not, however, cover their fees, and thus the fees were not justified. Finally, he concluded that for the average investor, an investment into these funds was not going to be worth it.

In an extensive study performed by Daniel, Grinblatt, Titman, and Wermers (1997), 2,500 U.S. mutual funds were investigated. The primary goal of the study was to uncover whether fees were justified by the stock-picking skill of mutual fund managers. Daniel et al. (1997) investigated the connection between funds excess return and characteristic selectivity and characteristic timing. From these connections, Daniel et al. (1997) found evidence of stock-picking abilities among the mutual fund managers. More specifically, this was found in aggressive-growth funds, with annual positive alpha. But, the alpha generated by these funds was very close to the fees collected by the funds, which, in turn, makes the performance of the fund from an investor's perspective neutral. Another extensive study on mutual funds was conducted by Blake and Timmermann (1998). With a sample of 2,300 U.K. mutual funds, they discovered that on a risk-adjusted basis, the average U.K. mutual fund in the period underperformed by about 1.8 percent.

Otten and Bams (2002) conducted a study where a total of 506 mutual funds from the

European countries Germany, France, Italy, Netherlands, and the U.K. were investigated. Funds that had been closed within the period of investigation were not eliminated from the sample. This was done to control for survivorship bias. In their paper, Otten and Bams (2002) used the four-factor model of Carhart (1997) in order to evaluate funds. From the results, they gathered evidence of positive alpha for returns where management fees had not been subtracted from the funds in all five countries except Germany. Otten and Bams (2002) also noted that the small-cap funds are especially able to add value for investors.

Edelen (1999) conducted a study where 166 U.S. mutual funds were under investigation. Edelen (1999) applied the single-factor model proposed by Jensen (1968), using the CRSP value-weighted index. Findings indicated that managers do little for investors besides collecting fees. Finally, Edelen (1999) concluded that the managers did not generate the negative alpha due to a lack of stock-picking skill, but rather as a result of the fees the managers collect.

The findings in studies presented so far have not made any clear cut answers as to whether mutual fund managers are able to outperform the market, and as such inhabit stock-picking skill. Some researchers find evidence of outperformance at least in the gross of fees returns. However, looking at the aggregate level, most studies have found evidence of zero or negative alpha. But, as argued by Grossman and Stiglitz (1980), due to momentary mispricing of market securities, some mutual funds outperform, and some underperform. Lastly, we note the importance of appropriate benchmarks in the studies.

2.2 Evidence from Scandinavia

Studies conducted on Scandinavian mutual funds are far fewer than those published regarding U.S. or larger countries in Europe. Yet, we have some evidence from Scandinavian markets in relation to mutual fund performance. Some of the studies have also investigated the subject of stock-picking skill vs. luck among the mutual funds investigated.

Sørensen (2009), as mentioned, conducted a study using the bootstrap procedure of Kosowski et al. (2006). Sørensen employed the modifications introduced in Fama and French (2010) to evaluate the performance of funds over a period from 1982 to 2008. Sørensen (2009), as researchers before him, ensured a sample free of survivorship bias by including funds that had been closed throughout the period. Through the findings, Sørensen (2009) concluded that there was no evidence of stock-picking skill among the top-performing funds. When investigating the worst performing funds of his sample, however, there was evidence of managers skill level leading to abnormal negative performance. Gallefoss et al. (2015) also examined Norwegian mutual funds through the methodology of Kosowski et al. (2006). Using daily return observations in the period 2000-2010, findings from their investigation suggested that there was significant evidence of both abnormal outperformance and underperformance among the Norwegian mutual fund managers. The performance was attributed due to skill or a lack thereof and not luck alone. However, when Gallefoss et al. (2015) turned to the aggregate level, funds were concluded to underperform relative to the benchmark.

In the Danish mutual fund market, Christensen (2003) investigated mutual funds from October 1994 to January 2002. Findings suggested neutral performance by most of the mutual funds and some negative. There was, however, some evidence of market timing skill amongst the mutual fund managers, yet most of these managers seemed to lack an ability in stock selection. Later on, Christensen (2013) investigated the returns of funds over the period 2000 to 2010. From the 71 funds investigated, a small fraction generated statistically significant alpha. However, the majority of funds generated negative alpha, where almost half of these were statistically significant. Christensen (2013) also included a discussion about the fees investors pay for entering and leaving the funds. Inclusion of these fees into the net-of-fees returns would decrease the funds generated alpha even further.

Dahlquist, Engström, and Söderlind (2000) conducted a study of the Swedish mutual funds in 1993-1997, where both performance and characteristics was investigated. Findings suggested that most funds managed to generate a neutral performance. Interestingly, regular equity funds were found to have an outperformance. From this, Dahlquist et al. (2000) concluded that their discoveries suggested that actively managed funds performed better than investing in a passive (index) fund. Flam and Vestman (2014) investigated actively managed Swedish equity mutual funds in the period between 1999-2009. Before expenses, the average four-factor alpha generated by funds was 0.9%. When expenses were deducted, however, the alpha turned negative, more specifically to -0.5 %. Findings suggested further that persistence in the returns was practically non-existent, and fund returns converged to the mean in a time span of approximately two years. In addition, Flam and Vestman (2014) investigated whether there was any evidence of skill, employing the bootstrap methodology of Kosowski et al. (2006) and Fama and French (2010). From their analysis, they found no evidence of skill amongst the mutual fund managers. In a later study, Flam and Vestman (2017) investigated a longer period of Swedish mutual fund returns. Here, they found evidence of outperformance in the Swedish mutual funds for the period of 1993-2001, but not in the other part of the study, 2002-2013. In this study, they employed a similar investigation as they did for the 1999-2009 sample, finding similar results.

Finally, the Finnish mutual fund market between 1993 and 2000 was studied by Korkeamaki and Smythe Jr. (2004). From their investigation, they found that the funds generally had a neutral performance. However, contradicting the findings in the Swedish mutual fund market, the equity funds mainly provided a negative performance relative to the benchmark.

The studies from Scandinavia offers contradicting evidence regarding both the performance of mutual funds and whether this is due to luck or skill in the respective countries. Therefore, we will dive deeper into the subject of luck versus skill in stock-picking among mutual fund managers. Several international studies have been conducted on the topic, some of which we will review in the following section.

2.3 Studies in Luck vs. Skill

Kosowski et al. (2006) apply a bootstrap method that is used to distinguish between skill and luck. As such, this methodology will be applied to the Norwegian mutual fund returns to determine whether the results of the mutual fund managers are due to skill or luck. The bootstrapping methodology offers an advantage in that one does not need to specify the shape of the distribution from which the returns are drawn (Kosowski et al., 2006).

White (2000) proposed a reality check in order to control for data snooping bias. This was further applied by Kosowski et al. (2006). From their investigation of U.S. mutual funds, Kosowski et al. (2006) found evidence of stock-picking skill, and lack thereof, for the top and bottom 10 percent of funds respectively. That is, the top funds yielded returns so good that it could not be attributed to luck alone, and, as such, had to be due to stock-picking skill. Whereas, the bottom performers generated returns so low that one could not blame it on bad luck alone, and some of the returns had to be attributed to the lack of skill in managers stock-picking.

As argued by Lo and MacKinlay (1990), data snooping occurs because "the more scrutiny a collection of data is subject to, the more likely will interesting (spurious) patterns emerge".

Based on this, it is argued that data snooping (mining) biases increase the more one studies the topic. Regarding the topic of luck vs. skill in the cross-section of alpha distributions, the chance of one finding abnormalities or patterns in the data is relatively high.

The paper presented by Yan and Zheng (2017) focuses on fundamental-based variables that were derived from financial statements. Yan and Zheng (2017) applied a bootstrap approach in order to evaluate the data-mining bias inherent in the cross-sectional return anomalies and to examine the fundamental signals derived from financial statements. Even after the application of the bootstrap method to account for data mining, Yan and Zheng (2017) found evidence that there are in fact elements that statistically significantly predict cross-sectional stock returns. The evidence points to a more pronounced predictive ability for small, high-volatility stocks, and that fundamental-based anomalies are based on expectations and mispricing as opposed to data mining.

Barras, Scaillet, and Wermers (2010) employed the method of Kosowski et al. (2006) in their attempt to separate fund performances based on luck from those performances based on skill. In addition to the bootstrapping methodology, Barras et al. (2010) suggested that one could use False Discovery Rates (FDR) in order to separate luck from skill. New FDR measures were developed. These new measures allowed them to individually target the top and bottom tails of the cross-sectional alpha distribution, and measured the impact of luck on the funds' performance. The findings of Barras et al. (2010) from US equity funds, indicated that 7.1 percent of managers were skilled through the standard approach. However, when accounting for luck, the funds were no longer able to generate positive alpha.

Fama and French (2010) modified the method proposed by Kosowski et al. (2006). Through their investigation of U.S. mutual funds in the period 1984-2006, they found contradicting evidence to Kosowski et al. (2006). Fama and French (2010) found no evidence of outperformance on a net of fees basis. They did, however, mention that findings suggested that at very high percentiles of the cross-sectional distribution of t-statistics of alpha, that funds gross of fees could outperform the market. Further, Fama and French (2010) found similar evidence to Kosowski et al. (2006) in the lowest percentiles, in that the bottom funds indeed suffer from a lack of skill in the management of these funds.

Recently, Harvey and Liu (2020) investigated the error rates of the Fama and French (2010) approach, using their proposed double bootstrap approach. From their investigation, Harvey and Liu (2020) found that the test lacks power in detecting outperformance. Through their study, they find that the test power increases when the minimum amount of observations is increased. When short series of returns are introduced into the bootstrapping methodology of Fama and French (2010), Harvey and Liu (2020) argue that the returns from the bootstrapping might be inflated, and as such, it is harder for a fund to overcome the statistical threshold and be found as skillful. When increasing the minimum length of the return series, Harvey and Liu (2020) found evidence of skill using a similar sample to Fama and French (2010). Note that when increasing the minimum amount of observations, some funds had to be excluded from the sample.

The method of Kosowski et al. (2006) was also used in a study conducted by Cuthbertson, Nitzsche, and O'Sullivan (2008) in the investigation of U.K. mutual funds in the period 1975-2002. Their findings supported those of Kosowski et al. (2006) in that bottom funds inhabited a lack of stock-picking skill, and the top performers had stock-picking abilities.

2.4 Test Power

Harvey and Liu (2020) recently introduced a new methodology where one can evaluate the Type I and Type II error rates of a statistical threshold or procedure using the data at hand. Through a so-called double bootstrap approach, Harvey and Liu (2020) argue that, building on the framework of False Discovery Rates, one is able to find empirical error rates for statistical thresholds when applied to the sample at hand.

When the Type I error rate is high, this means that we have a high probability of falsely rejecting the null hypothesis. If we are in a situation where the Type II error is high, this implies that we have a high probability of failing to reject the null when the alternative hypothesis is true. From a high Type II error, it follows that the test power is low (test power = 1 - Type II error rate).

Ioannidis (2005) states that a low test power for alternative hypotheses, in turn, generates a high Type I error rate for the main hypothesis tested. Andrikogannopoulou and Papakonstantinou (2019), Ferson and Yong (2017), and Barras, Scaillet, and Wermers (2019) have also conducted studies with focus on the FDR approach of Barras et al. (2010), which investigated skill or luck among US equity funds. Further, researchers has developed multiple testing techniques in related fields in financial economics. Harvey, Liu, and Zhu (2016) focus on the Familywise Error Rate (FWER). In their paper, they present a stylized bayesian framework for multiple testing. Through that approach, one indirectly achieves the multiple-testing adjustment through the likelihood function.

Harvey (2017) suggests a bayesianized p-value, where the p-value of a test is corrected. Here, one applies the minimum bayes factor. The minimum bayes factor adds to the bayesian hypothesis testing but separates itself from the prior specification in that it focuses on the prior that generates the minimum bayes factor. Barras et al. (2019) employs an FDR approach with corrections. Despite the number of methods developed, we still face an obstacle as to which correction method is most suitable for a given dataset. This cuts to the core of the approach proposed by Harvey and Liu (2020), where the multiple testing correction is conducted via a systematic approach, offering data-driven guidance on the relative performance of multiple testing and adjustment methods.

The treatment of p_0 in Harvey and Liu (2020) lies between the approaches of Harvey et al. (2016) and Harvey (2017). Meaning that while the new approach does not delve into the process of making assumptions on the prior distribution of p_0, which one would later feed into the proposed full-blown Bayesian framework (as is done in Harvey et al. (2016)), we also diverge from the Harvey (2017) assumption of a degenerate prior. That is, the point mass which concentrates on the value of the parameter that generates the minimum bayes factor. This is done by exploring how the Type I and Type II error rates act as a response to changes in p_0.

Building on Benjamini and Hochberg (1995), Genovese and Wasserman (2002) and Sarkar (2006) proposed an alternative way to measure the Type II error rate, the false non-discovery rate. The false non-discovery rate is defined as the portion of false non-rejections among those tests whose null hypotheses are not rejected. Here, Harvey and Liu (2020) rather applies a measure where their definition of the Type II error rate is the number of false non-rejections divided by the number of true negatives plus false negatives. This is done to differentiate between cases where the number of tests are very different. This is further explained in Section 3.2.3.

3 Methodology

This section aims to provide readers with the models applied in this thesis to investigate the performance of mutual fund managers. In addition, the method of bootstrapping will be discussed, along with how this can be used to distinguish skill and luck. As evident from many studies, it is within reason to expect abnormal performance in the tails of the distribution. Therefore, separating skill from luck will be crucial. In order to distinguish this, the bootstrap approach of Kosowski et al. (2006), Fama and French (2010), and Harvey and Liu (2020) is applied.

3.1 Model Selection

The well-known abnormal performance measure, alpha, will be used in order to evaluate the performance of mutual funds. Funds are ranked by their t-statistic of alpha to take into consideration the sampling uncertainty. Through a time series regression in R, the abnormal performance of funds excess return is estimated.

3.1.1 Single-factor Model

The previous literature review introduced some models used in past studies. Here, a more detailed description of Jensen's alpha is provided, which is an essential building block for the models we will discuss. Jensen (1968) based the single-factor model on the Capital Asset Pricing Model, revealing the relationship between risk and return for assets, based on the exposure of the asset to the market factor. The CAPM was altered by Jensen, adding Jensen's alpha, in order to measure whether an asset would create an abnormal return according to CAPM. If we observe a positive alpha, this states an outperformance in relation to CAPM. Conversely, a negative alpha correspond to an underperfomance. If markets are, as proposed by Fama (1970), perfectly efficient, alpha would disappear, equalling zero. The single-factor model is stated as

$$r_{j,t} = \alpha_j + \beta_j MKT_t + \epsilon_t, \tag{1}$$

where, $r_{j,t}$ is the excess return of asset j in time period t, α_j expresses the abnormal return of asset j, and β_j represents the change in the assets excess return in relation to the markets excess return. That is, the beta measures the sensitivity of the asset in relation to the market.

MKT is the market excess return in the same period. Finally, ϵ_t represents the individual assets specific risk. This is the non-systematic risk, which is diversifiable.

3.1.2 Fama-French Three-factor Model

The single-factor model proposed by Jensen (1968) was debated in the '80s and '90s. Researchers (Reinganum, 1981; Breeden, Gibbons, & Litzenberger, 1989) argued that the model was not complete enough to explain the return of assets. Fama and French (1993) improved on the single-factor model of Jensen, arguing that additional risk factors contributed to returns. By adding the two factors of size (SMB) and value (HML), they constructed the three-factor model. The size factor that Fama and French (1993) created, was made by putting companies on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ Stock Market into different portfolios. The companies were selected for the different portfolios based on size. The same process was used for the value factor.

The two factors, SMB and HML, warrant some further explanation. SMB stands for Small (S) capitalization, Minus (M) the portfolio return of Big (B) capitalization companies. As for the HML-factor, accounting for the value premium, companies were put into one of two groups based on their book-to-market ratio, High(H) or Low (L), M represents a minus. Bauman, Conover, and Miller (1998) conducted a study where they argue that over time, small-capitalization companies generate higher returns than what large companies do, backing the HML factor of Fama and French (1993).

Grouping the companies were a continuous process throughout the years of the conducted research from 1963 up to 1993. From these groups, they then produced six portfolios in order to create the baseline for the SMB and HML factors, all represented in the equations below. To calculate the SMB-factor, we use the following equation

$$SMB = \left(\frac{1}{3}S/H + \frac{1}{3}S/M + \frac{1}{3}S/L\right) - \left(\frac{1}{3}B/H + \frac{1}{3}B/M + \frac{1}{3}B/L\right). \tag{2}$$

The idea behind the SMB-factor is that it enables the comparison of low market-value companies' returns to the companies with high market value. The following equation shows how the

HML-factor is calculated,

$$HML = \left(\frac{1}{2}S/H + \frac{1}{2}B/H\right) - \left(\frac{1}{2}S/L + \frac{1}{2}B/L\right), \tag{3}$$

HML represents the average return of high book-to-market ratio minus low book-to-market ratio average returns, i.e., we are subtracting the return of growth portfolios from the value portfolios. From equation 3, we gather that a positive HML value corresponds to a higher return in value portfolios than that of growth portfolios, and a negative HML value correspond to a higher return in growth portfolios relative to value portfolios.

With the two factors in equation (2) and equation (3) combined with the model (1), we can generate the proposed Fama-French three-factor model:

$$r_{j,t} = \alpha_j + \beta_j MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{j,t}, \tag{4}$$

where the coefficients β_2 and β_3 represent the effect of the SMB-factor and HML-factor in period t on the excess return of the asset.

3.1.3 Carhart Four-factor Model

The three-factor model proposed by Fama and French (1993) was further extended by Carhart (1997). Carhart added a momentum factor (PR1YR), based on the study of Jegadeesh and Titman (1993), which aimed to incorporate the one-year momentum anomaly. The one-year momentum factor accounts for the market anomaly that increasing (decreasing) stocks will continue to increase (decrease) in the following month. In order to generate the momentum factor, a portfolio of the best performing stock returns is subtracted the return of a portfolio of the worst-performing stocks. Adding the momentum factor to the three-factor model of Fama and French results in the following model:

$$r_{j,t} = \alpha_j + \beta_j MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 PR1YR_t + \epsilon_{j,t}, \tag{5}$$

in equation (5) we see that the additional coefficient, β_4, represents the effect of the one-year momentum factor on the return of the asset. The rest of the coefficients and risk factors correspond to the explanations to model (1) and model (4). The four risk factors in Carharts model are not diversifiable. The four-factor model will be the main performance model in the study, as it is used by Kosowski et al. (2006) and Fama and French (2010). Further, it will be applied in the multiple test proposed by Harvey and Liu (2020).

That said, both the single-factor and three-factor models will be applied to illustrate differences in results.

3.2 Bootstrapping

Bootstrapping is used in the methodology corresponding to Kosowski et al. (2006) and Fama and French (2010) in separating skill from luck. In addition, the methodology of Harvey and Liu (2020), accounting for error rates applies bootstrapping. The use of bootstrapping circumvents the problems of parametric tests when dealing with intricate dependencies in the cross-section of non-normal returns. The joint-test bootstrap procedure relies on fewer assumptions due to its robustness to dependencies in the cross-section. Also, it allows for general distributional characteristics in that one does not have to assume any distributional characteristics. Lastly, the procedure automatically takes the sampling uncertainty into account (Yan & Zheng, 2017).

When applying the bootstrap methodology, one also controls for data snooping. Data snooping as presented by, as mentioned, White (2000) is also known as data mining. Data snooping is a pitfall when conducting statistical analysis. In this thesis, data snooping could occur, as we estimate the same measures for multiple entities. As such, findings from individual Carhart four-factor regressions might be biased. Another example could be that one is looking for correlations in a massive data sample, and only report a small subset of this. Then, as White (2000) presented it, data snooping occurs. Further, this can result in forecasting models that look usable, despite there not being any exploitable forecasting relations in the sample, due to studying the sample at lengths. To control for this, White (2000) applied a Bootstrap Reality Check, referred to as White's Reality Check.

Bootstrapping procedures can be used to estimate the unknown distribution of a parameter. This is why the bootstrapping procedure is so appealing in an investigation of mutual funds. We are here able to investigate skill vs. luck by generating a distribution of the alpha under the null of no skill. That is, if we set alpha to zero, what would the distribution of alpha be? From this, we can infer whether the generated alpha of a fund is due to luck or skill.

The bootstrapping procedure is a cross-sectional methodology. In traditional Ordinary

Least Square (OLS) regressions, a key assumption is the normal distribution of residuals. Several properties within the data may lead to this assumption not holding. Statistical significance is easier to test in the framework of Jensen's alpha in comparison to the Sharpe ratio. This is due to the fact that Share-ratio fails to differentiate anomalies, leading to the ability to manipulate results in the use of the Sharpe ratio. The use of factor models is, however, not a flawless approach, as the separation of luck and skill is not always possible. Both the cross-section of mutual fund alpha and the individual fund level alpha will be studied on possible non-normalities.

The reason why fund returns experience kurtosis and skewness not corresponding to the normal distribution since fund managers usually hold big positions in a lower amount of stocks, rather than the opposite. As such, the central limit theorem fails, i.e., we do not experience the phenomena where an equally weighted portfolio of non-normally distributed stock returns approach normality. This fact, in turn, invalidates the use of parametric tests. There is also a tendency for stock returns to be heteroscedastic, and, in addition, have autocorrelation in the series.

The last point concerns the investment strategies of the mutual funds. The different investment strategies may lead to different and altering risk preferences. Altering risk preferences may be due to performance compared to similar portfolios or funds, or the overall market (Kosowski et al., 2006). When the mentioned properties affect the sample, non-normality in alpha distribution occurs. Applying the Jarque-Bera test, one can evaluate the normality of residuals in the models presented above. For instance, when applying the Jarque-Berra test to the estimation of the Carhart (1997) four-factor model of equation (5), normality in mutual fund residuals is rejected frequently.

This is where the bootstrapping methodology comes in handy. As we circumvent distributional assumptions, the validity of inferences may be improved significantly. Researchers have widely argued that bootstrapping enables a more definite assessment of alpha estimate significance (Bickel & Freedman, 1984; Fama & French, 2010; Hall & Martin, 1988; Horowitz, 2003; Kosowski et al., 2006). Through a Monte Carlo experiment, Horowitz (2003) were able to illustrate the effect of bootstrapping. His findings suggested that there was a significant reduction between the true probability and the nominal probability of correctly rejecting the null, i.e., reduction of the Type I Error rate. Through Horowitz (2003) study, he proved that the standard parametric t-test rejects abnormal performance less frequently than the bootstrapping procedure. The key being that bootstrapping procedures recognize the thicker than normal tails often found in fund returns.

Through the cross-sectional distribution of mutual fund residuals being built up of individual fund distributions, and the assumption of normalities is disabled. These funds are identified by: 1. Heterogeneous risk-taking in funds throughout the sample and 2. increased momentum in individual fund residuals. Despite funds having normally distributed residuals, this does not automatically lead to the cross-sectional distribution being normal. When considering the joint distribution of all funds in the sample, the bootstrapping approach will allow for more accurate inferences than if we were to rely on parametric assumptions and conducting a parametric test. The bootstrapping is done by resampling with replacements, and through this generating the statistic for each resample.

3.2.1 Kosowski et al. Approach

The Kosowski et al. (2006) bootstrapping procedure conducts its procedure under the null hypothesis of zero true performance, or equivalently, no superior fund manager exists. The alternative hypothesis of this is that we have true performance in the tails. Specifically, that $\alpha_{Top} > 0$ and $\alpha_{Bottom} < 0$. The basis for the bootstrap procedure of Kosowski et al. (2006) is created through the estimation of the OLS regression of the Carhart (1997) four-factor model presented in equation (5). We estimate the four-factor model for each individual fund, j, ending up with the following estimated model,

$$\hat{r}_{j,t} = \hat{\alpha}_j + \hat{\beta}_{1,j} MKT_t + \hat{\beta}_{2,j} SMB_t + \hat{\beta}_{3,j} HML_t + \hat{\beta}_{4,j} PR1YR_t + \hat{\epsilon}_{j,t}. \tag{6}$$

From the OLS regression, one needs to store the coefficient estimates for each fund j, $\hat{\alpha}_j, \hat{\beta}_{1,j}, \hat{\beta}_{2,j}, \hat{\beta}_{3,j}, \hat{\beta}_{4,j}$. In addition, the estimated residuals for each time period and each fund are stored, $\hat{\epsilon}_{j,t}$, where t goes from the first to the last month of observation of fund j respectively. Lastly, the t-statistic of alpha for each fund is stored, $\hat{t}_{\hat{\alpha}_j}$. One could equivalently to storing the coefficient estimates store the fitted values from the regression, and apply these in order to generate the pseudo sample of returns below instead of coefficient estimates.

From the stored regression estimates, one turns to the construction of a pseudo-random time-series of resampled residuals. This is created through randomly drawing samples, with replacement from the estimated residuals of fund j. We impose the restriction of the pseudo sample having the same length as the original return series. The length is given by, $\hat{\epsilon}^b_{j,t_\epsilon}, t_\epsilon = S^b_{t_j,0,....,T_{j,1}}$. Here, b represents the bootstrap iteration, whereas $t, ..., T$ defines the starting point and end of the series. After having created the pseudo-sample of residuals, we fetch the saved coefficient estimates.

Through combining the pseudo-sample of residuals and the coefficient estimates, a pseudo-monthly time-series of excess returns for fund j is created. In this step, the null hypothesis of zero alpha performance is implemented. This is done via subtracting the estimated alpha from each fund's returns. Due to sampling uncertainty, this will not result in zero alpha for the pseudo-monthly time-series, although the null will still hold. We estimate the four-factor model of equation (6) for the pseudo sample of excess returns for each fund. The pseudo sample of excess return is given by

$$r^b_{i,j} = 0 + \hat{\beta}_{1,j}MKT_t + \hat{\beta}_{2,j}SMB_t + \hat{\beta}_{3,j}HML_t + \hat{\beta}_{4,j}PR1YR_t + \hat{\epsilon}^b_{i,t}, \tag{7}$$

where the pseudo sample return has zero alpha, and consists of multiplying the estimated coefficients of the fund to the risk factors in a given month. We add the bootstrapped residual, and then proceed to run another OLS regression on the pseudo sample return series. The procedure will be conducted $J = 10,000$ times for each individual fund. As such, we may obtain a non-zero estimated alpha for a random bootstrap iteration b. That is, when we apply the four-factor model to the pseudo return series, it may occur that for a given bootstrap sample, we have drawn an abnormal amount of positive or negative alpha, resulting from the sampling variation surrounding the zero true performance and the drawn residuals.

For each run of the 10,000 bootstrap iterations, we build the cross-sectional distribution of bootstrapped t-statistic of alpha for each fund ($i = 1, 2, ..., 107$). We thus end up with 10,000 cross-sectional distributions of bootstrapped t-statistics. We place all bootstrapped t-statistics into a matrix, creating a $J \times N$ matrix, where J is the number of bootstrap iterations, and N is the number of funds. For each row in the matrix of bootstrapped t-statistic, we sort the t-statistics from highest to lowest. That is, the highest bootstrapped t-statistic is found in column 1, and the lowest is found in column N, for each bootstrap iteration.

Finally, the bootstrapped p-values is calculated as the sum of bootstrapped t-statistics greater than the actual t-statistic divided by the number of iterations when the actual t-statistic is greater than 0. When the actual t-statistic is smaller than 0, the bootstrapped p-value is calculated as the sum of bootstrapped t-statistics lower than the actual t-statistic divided by the number of iterations. More specifically, for each fund we apply the following equation,

$$\text{Bootstrapped } p - value_i = \begin{cases} \frac{\sum_{j=1}^{J} t_{i,bootstrap} > t_{i,actual}}{J}, & \text{if } t_{i,actual} > 0 \\ \frac{\sum_{j=1}^{J} t_{i,bootstrap} < t_{i,actual}}{J}, & \text{if} t_{i,actual} < 0, \end{cases} \tag{8}$$

where $t_{i,bootstrap}$ represent the bootstrapped t-statistics at percentile i and $t_{i,actual}$ represent the actual t-statistic at percentile i from the original data.

The reason for using the t-statistic of alpha, as opposed to the alpha estimate itself, is due to the fact that the t-statistic of alpha has been argued to have a superior predictive ability as a sorting term when comparing performance under the assumption of zero alpha across all funds (Busse, Goyal, & Wahal, 2010; Fama & French, 2010; Kosowski et al., 2006). This is because mutual funds with short return series characterized by high risk will have greater variance in estimated alpha distribution, which, in turn, generates outliers in the cross-section. The t-statistic of alpha addresses this issue and penalizes a high alpha estimate connected to high variance by reducing the corresponding t-statistic through the standard error.

The bootstrap approach of Kosowski et al. (2006) can be summarized in the following steps:

I Estimate the model in equation (6) and store the estimates of alpha, factor loadings, residuals, and t-statistic of alpha for each fund.

II Create a pseudo return series, as in equation (7) for each individual fund, with zero alpha imposed, through bootstrapping the residuals of the fund.

III Regress the pseudo return series on factor returns, store the bootstrapped t-statistic of alpha for the pseudo return series.

IV Repeat steps II and III J times for each fund i in the sample.

V Sort each row of the $J \times N$ matrix of bootstrapped t-statistics from low to high. Making

J CDFs of bootstrapped t-statistics.

VI Calculate the bootstrapped p-value for each percentile as in equation 8.

3.2.2 Fama & French Approach

Next, we turn to the approach of Fama and French (2010), where the goal of the approach is to draw inferences regarding the cross-section of the true α for actively managed funds. Specifically, if we find evidence from the cross-section of alpha estimates that suggest a world with true α of zero for all funds, or whether we, in fact, experience nonzero true α. We especially regard the tails of the cross-section of α estimates. In this thesis, the net returns of the Norwegian mutual funds are applied. When setting $\alpha = 0$ for the net returns of the funds, the null hypothesis is the following: We live in a world where mutual fund managers can generate returns covering the fees they require. The alternative is thus that performance in the tails is due to skill as opposed to luck. Specifically, that $\alpha_{top} > 0$ for the top performers and $\alpha_{bottom} < 0$ for bottom performers.

Fama and French (2010) and Kosowski et al. (2006) frameworks are pretty similar, however, there are differences that makes applications of both in investigation interesting. Kosowski et al. (2006) imposes their null hypothesis of zero alpha through subtracting estimated alpha from fitted values. Fama and French (2010) impose the null through subtracting the alpha estimate from the actual return series of funds. In Fama and French (2010) we simulate fund returns and factor returns jointly, whereas Kosowski et al. (2006) bootstrap residuals only in their baseline bootstrap.

Preceding the bootstrapping procedure, the setup is conducted in the following way. We run an OLS regression, as in the Kosowski et al. (2006) framework, estimating the same model of equation (6), and storing the same estimates. The following steps are where the framework of Fama and French (2010) differs from that of Kosowski et al. (2006).

In the Fama and French (2010) test for nonzero true α in the actual fund returns, one bootstrap in order to generate return series which have the property of fund returns. The difference, true α is set to zero for each and every fund. Specifically, we subtract the estimated alpha from the excess returns of the respective funds ($r_{i,t} - \hat{\alpha}_i$). From this we estimate, fund by fund, the statistic of the bootstrap procedure, using a random sample with replacement of the 396 months drawn from the full length of the time series sample of 396 months.

When conducting the Kosowski et al. (2006) approach, the random sample of monthly returns is different for each fund in each bootstrap iteration. Whereas in the bootstrapping approach of Fama and French (2010), it is done with the same random sample of months for all funds. That is, in the Kosowski et al. (2006) approach, the first observation for fund j might be $t = 10$, whereas the first observation of fund k might be $t = 96$ in the same bootstrap iteration. For the Fama and French (2010) approach, it is ensured that the $t = 10$ is used in the first observation of all funds in a given iteration. This ensures capturing cross-correlation of the fund returns, and the effects on the distribution on the t-statistic of alpha estimates.

The joint sampling approach of sample fund and explanatory returns captures any correlated heteroscedasticity in the explanatory returns and disturbances in the benchmark model. However, this also leads to a loss of any potential effect of autocorrelation in the factor returns, and losing any effects of variation through time in the regression slopes. Using the bootstrapped time series, we construct the following excess return series:

$$(r_{i,t} - \hat{\alpha}_i) = \hat{\beta}_{1,i} MKT^b + \hat{\beta}_{2,i} SMB^b + \hat{\beta}_{3,i} HML^b + \hat{\beta}_{1,i} PR1YR^b + \tilde{\epsilon}_i^b. \tag{9}$$

From the pseudo return series generated above, we regress, on the risk factors for each fund. From the regressions, we then store the estimated t-statistic of alpha of the pseudo excess return series.

After having run $J = 10.000$ bootstrap simulations, we estimate the t-statistics of alpha for the pseudo excess returns. We then end up with $J = 10.000$ cross-sectional distributions of bootstrapped t-statistics of alpha. As in the Kosowski et al. (2006) approach. We next sort each of the cross-sectional distributions from high to low.

Fama and French (2010) draw inferences in a similar manner to that of Kosowski et al. (2006). However, they also compare the average bootstrapped t-statistics over the 10,000 bootstrap iterations at a given percentile to the actual t-statistic at the same percentile drawn from the actual t-statistic distribution. In addition, rather than presenting the bootstrapped p-values, they present likelihoods.

The likelihoods are calculated by finding the percent of bootstrapped t-statistics at a given percentile that is below that of the same percentile in the actual t-statistics distribution. If we then experience a low likelihood for the worst performers, we can infer that some managers in fact lack skills that suffice to cover their fees. For ease of comparison for the two approaches, bootstrapped p-values are calculated from the bootstrapping procedure of Fama and French (2010).

The Fama and French (2010) approach can be summarized in the following steps:

I Estimate the model in equation (5) and store the estimates of alpha, factor loadings, residuals, and t-statistic of alpha for each fund.

II Generate pseudo excess return series with imposed zero alpha for each fund, as in equation (9), by bootstrapping factor returns and residuals through a common time ordering for all funds in a given iteration.

III Regress the pseudo return from step II on the factor returns. Store the bootstrapped t-statistic of alpha for each fund.

IV Repeat steps II-III J times, in order to generate J cross-sectional distributions of bootstrapped t-statistic of alpha.

V Sort each row in the $J \times N$ number of funds matrix from low to high.

VI Calculate the bootstrapped p-value for each percentile as in equation 8.

3.2.3 Double Bootstrap Approach by Harvey & Liu (2020)

Harvey and Liu (2020) also apply bootstrapping in their paper, although they first and foremost apply the bootstrapping methodology to evaluate error rates of statistical thresholds or statistical procedures.

Harvey and Liu (2020) argue that by running two rounds of bootstrapping, both rounds based on the original data, one can evaluate error rates of statistical thresholds specific to the data set at hand. The reason why this approach is so appealing is that the Type I errors are likely too frequent if one applies single-hypothesis testing criteria in a multiple hypothesis situation. Due to the high Type I error rate follows a large number of managers will look like they outperform the market purely by luck.

The double bootstrap approach first determines a performance threshold that delivers a particular Type I error rate. Next, one characterizes the Type II error rate associated with this optimized Type I error rate.

In the framework of Harvey and Liu (2020), one assumes that a fraction of mutual fund managers p_0 has skill, or a lack thereof if one wishes to evaluate the bottom performers. The skill level of managers is set at the in-sample performance measurement estimate. The rest of the managers, $1 - p_0$, are assumed to have no skill, setting their skill level to 0 (zero excess return or zero alpha).

When bootstrapping from these adjusted return series, one is able to evaluate the Type II error rate through the use of simulations. Through the use of p_0, the difficulty in setting the high-dimensional parameter vector for the alternative hypothesis is circumvented. The parameter vector is set to a reasonable value, e.g., the in-sample estimate corresponding to a certain p_0. In essence, we treat p_0 as a sufficient statistic, helping in the estimation of the Type II error rate.

p_0 has several interpretations. When p_0 is set to zero, no manager is believed to be in possession of skill. This correspond to the null hypothesis applied in Kosowski et al. (2006) and Fama and French (2010). Setting p_0 greater than zero, assume that some managers are believed to be outperforming. In this case, one can think of p_0 as a plug-in parameter, similar to the role of μ_0 in a single test of the mean of a statistic, which helps measure the error rates in the presence of a multiple test.

In running the double bootstrap, one selects p_0 outperformers in the sample, which are, in the second round bootstrap, known as true outperformers. The reason for using two rounds of bootstrapping is that if one instead chose the p_0 true strategies from the actual ranked t-statistics, one would be ignoring the sampling uncertainty. When bootstrapping the data in an initial first-round bootstrap, and then selecting the top-performing funds based on the highest t-statistics, the sampling uncertainty is circumvented.

The double-bootstrap method starts by selecting $p_0 \times N$ funds which are deemed to be truly outperforming. To take sampling uncertainty into account, one perturbs the data, and then ranks the funds based on the t-statistics from the perturbation. Specifically, the periods of the original data X_0 are bootstrapped, giving an alternative panel of returns, X_i. Next, for the panel X_i, funds are ranked based on their t-statistics.

For the top p_0 funds, with the highest t-statistics, we locate the funds in the original sample, X_0. The corresponding funds are adjusted to have the same alpha as that creating the t-statistic in X_i. The data matrix of adjusted funds having non-zero alpha is denoted $X_{0,1}^{(i)}$. The remaining funds, $1-p_0$, are adjusted to have zero abnormal performance, i.e. alpha set to zero. We denote the matrix of zero-alpha funds $X_{0,0}^{(i)}$. Lastly, $X_{0,1}^{(i)}$ and $X_{0,0}^{(i)}$ are arranged into one new data panel, Y_i. The return series in Y_i is hypothetical data, which is used to evaluate the follow-up error rate analysis, where the true number of over- /underperformers is known.

The bootstrapping of the adjusted panel of returns Y_i allows us to evaluate error rates, as mentioned, for a statistical procedure. This will be done for fixed t-statistic thresholds. Via the creation of the two adjusted panels and combining them into one, we know exactly which funds in the panel are actually truly outperforming, and which are falsely outperforming.

By doing this, we can summarize the testing outcome of a statistical threshold for the j^{th} bootstrap iteration. We place these outcomes into a vector, $\bar{O}^{i,j} = (TN^{i,j}, FP^{i,j}, FN^{i,j}, TP^{i,j}$, where $TN^{i,j}$ (True Negative) corresponds to the number of tests that correctly identify a zero-alpha fund as having zero alpha. $FP^{i,j}$ (False Positive) is the number of zero-alpha funds identified as being outperforming, $FN^{i,j}$ (False Negative) is the number of tests that incorrectly identify a true outperformer as not outperforming. Finally, $TP^{i,j}$ (True Positive) is the number of tests that correctly identifies a true outperformer as outperforming. With the summary statistics in the vector $\bar{O}^{i,j}$, we are able to construct the two error rates used in the evaluation of the statistical thresholds.

The first, motivated by the False Discovery Rate (FDR), is the realized FDR. The realized FDR is defined as

$$RFDR^{i,j} = \begin{cases} \frac{FP^{i,j}}{FP^{i,j}+TP^{i,j}}, & \text{if } FP^{i,j} + TP^{i,j} > 0 \\ 0, & \text{if} FP^{i,j} + TP^{i,j} = 0 \end{cases} \tag{10}$$

That is, the realized FDR measures the fraction of false rejections of the null among all rejections of the null, both false and true. The Type I error rate in a single test is extended by the expected value of the realized FDR. The second error rate we measure is the realized rate of misses, also referred to as the false non-discovery rate or false omission rate. The realized rate of misses is also motivated by the FDR and is defined as

$$RMISS^{i,j} = \begin{cases} \frac{FN^{i,j}}{FN^{i,j}+TN^{i,j}}, & \text{if } FN^{i,j} + TN^{i,j} > 0 \\ 0, & \text{if } FN^{i,j} + TN^{i,j} = 0. \end{cases} \tag{11}$$

From this, we see that the fraction of missed rejections of the null (i.e., $FN^{i,j}$) among all the insignificant outcomes $FN^{i,j} + TN^{i,j}$ measures the realized rate of misses. This extends the Type II error rate in a single test through the expected value of RMISS.

Harvey and Liu (2020) also define a realized ratio of false discoveries to misses,

$$RRATIO^{i,j} = \begin{cases} \frac{FP^{i,j}}{FN^{i,j}}, & \text{if } FN^{i,j} > 0, \\ 0, & \text{if } FN^{i,j} = 0. \end{cases} \tag{12}$$

This is similar to the odds ratio concept used in Bayesian analysis. We here take the ratio of false discoveries, $FP^{i,j}$, to misses, $FN^{i,j}$. This ratio can help us in selecting the appropriate cutoff t-statistic based on one's belief of the cost of making Type I and Type II errors.

Finally, the sampling uncertainty due to ranking the funds and generating realized error rates for each of these rankings is accounted for by averaging across both i and j. Thus, if we suppose that we perturb the original data I times, and then for each i generate J bootstrapped random samples, we have:

$$TYPE1 = \frac{1}{IJ} \sum_{I}^{i=1} \sum_{J}^{j=1} RFDR^{i,j}. \tag{13}$$

Here, TYPE1 is referred to as the Type I error rate. Next, we define TYPE2 as

$$TYPE2 = \frac{1}{IJ} \sum_{I}^{i=1} \sum_{J}^{j=1} RMISS^{i,j}, \tag{14}$$

and refer to this as the Type II error rate. Finally, the sampling uncertainty is accounted for in the odds ratio

$$ORATIO = \frac{1}{IJ} \sum_{I}^{i=1} \sum_{J}^{j=1} RRATIO^{i,j}. \tag{15}$$

Note that the estimated Type I error rate and Type II error rate implicitly depend on the significance threshold applied in the calculations. Further, the Type II error rate is linked to power, and the RMISS in equation (11) definition implies a different interpretation of test power. Note that the way in which Harvey and Liu (2020) have defined the realized rate of misses implies that one takes into account the number of funds or strategies. When including all true discoveries into the denominator in the fraction for RMISS, it is implied that the statistical threshold is more impressive if it is possible to detect a small number of true outperformers in a large sample than detecting the same number of outperformers in a small sample.

The way in which the Type II error rate is usually, and most frequently calculated for multiple tests in other studies, extending the Type II error in a single hypothesis test is defined as:

$$TypeII^{i,j}_{usual} = \frac{FN^{i,j}}{FN^{i,j} + TP^{i,j}}. \tag{16}$$

As mentioned, this generates a different Type II error rate than the proposed calculation of Harvey and Liu (2020). Here, the sampling uncertainty mentioned above is taken into account in the same manner. More specifically, by

$$TYPEII_{usual} = \frac{1}{IJ} \sum_{I}^{i=1} \sum_{J}^{j=1} TYPEII^{i,j}, \tag{17}$$

we get the Type II error rate for the total simulations in the double bootstrap approach.

The Harvey and Liu (2020) double-bootstrapping methodology can be divided into a four-step procedure.

I Run the first round bootstrap. That is, bootstrap the time periods in the sample to obtain a new, bootstrapped return series matrix, X_i. Next, estimate the corresponding 1xN vector of t-statistics, t_i.

II Rank the t-statistics found in t_i. Then, find the corresponding funds in the original data matrix X_0 for the p_0 top funds. Next, adjust the corresponding strategies so that they have the same alpha as that generated by X_i. We denote the matrix containing adjusted fund returns $X^{(i)}_{0,1}$. The remaining funds in X_0 are adjusted to have zero alpha and collected in a matrix denoted $X^{(i)}_{0,0}$. Lastly, combine the two adjusted matrices $X^{(i)}_{0,1}$ and $X^{(i)}_{0,0}$ into one new panel of returns, Y_i.

III Bootstrap the time periods in Y_i J times. Then, for each bootstrapped sample, j in one to J, calculate the realized error rates for Y_i, which we denote by $f_{i,j}$ (f represents a generic error rate that is a function of the testing outcomes).

IV Repeat step I to step III I times. Finally, the one calculates the final bootstrapped error rate as:

$$\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} f_{i,j} \tag{18}$$

3.2.4 Bootstrap Extensions

The bootstrapping procedure for the basis of this thesis is a stationary bootstrap, as proposed by Politis and Romano (1994) by re-sampling the return residuals in data blocks. In the Fama and French (2010) framework, we allow for cross-correlation in the residuals and correlation in factor returns and residuals. Robustness is checked in the Kosowski et al. (2006) framework through varying block lengths of the bootstrap.

The distribution of bootstrapped t-statistics of alpha is smoothed by generating portfolios of funds. This allows for inspection of whether cross-sectional individual fund alpha analysis is affecting inference tests. Here, the corresponding average statistic in each tail of the portfolios is considered.

The lengths of data records are also considered. This is done by varying the minimum amount of monthly returns. Results of bootstrap extensions are presented and discussed in Section 5.3.

4 Data

This section aims to present the data used in the evaluation of the Norwegian mutual fund performance in the sample. The review of data furthermore creates the basis for empirical analysis presented in Section 5.

4.1 Norwegian Mutual Fund Sample

The dataset used in this thesis is comprised of 107 actively managed Norwegian mutual funds, which all have varying lengths of returns recorded in the time period of 1987 to 2019. Funds are required to have a minimum of 12 monthly returns in order to be included in the sample.

There is, however, no exclusion of funds that have closed throughout the period or opened after the inception of the period.

Funds are also required to hold at least 80% domestic equities to be eligible for the sample. Lastly, funds with neutral investment strategies, i.e., passively managed or index funds, are not considered. The choice of time period is based on the availability of sufficient market information on both fund and benchmark returns. Here, the restriction of only allowing Norwegian mutual funds into the sample is done to safeguard comparison of fund returns to an appropriate benchmark. The risk exposure varies from market to market, and the limitation on funds allows for the use of only one benchmark.

The TITLON database (Titlon, 2020) provides historical fund data, including complete daily information on the funds' Net Asset Value (NAV) throughout the period. The NAV of a fund represents its per-share market value. We use adjusted NAV, where ongoing costs such as management fees have been deducted, and the dividends are taken into consideration. From the last reported adjusted NAV of each month, returns for each fund have been created. The adjusted NAV of fund i in a given time period t provides the net of costs one-month simple return via the following equation:

$$r_{i,t} = \frac{NAV_t - NAV_{t-1}}{NAV_{t-1}}. \tag{19}$$

The average lifetime of funds in the sample is approximately 13.6 years. Table A.I., Appendix A, displays the number of observations for each fund in the thesis, along with various descriptive statistics of the funds' returns for the whole period the funds are open in the dataset.

All funds in the Norwegian fund market are domestic open-end funds. An open-end fund is where shares in the fund may be redeemed or issued without any limitation, as opposed to a closed-end fund, where investors wanting to sell their shares in the fund have to find a buyer willing to pay the price. As such, for closed-end funds, the price for a share in the fund may differ from the NAV of the fund.

4.2 Interest Rate

In order to calculate the excess return of both the individual funds and the benchmark, one needs to subtract the risk-free rate of return. In economic literature, treasury bills returns are widely accepted as a proxy for the risk-free rate of return. The liquidity of Norwegian treasuries is poor in comparison to other more comprehensive markets. Instead, the risk-free rate is constructed based on the one-month Norwegian Interbank Offered Rate (NIBOR) as argued by Ødegaard (2021b). The NIBOR aims to reflect the interest rate of unsecured money market lending between banks. Due to messy data in the years prior to 1987, for the one-month NIBOR, it would be difficult to incorporate this as the risk-free rate proxy for years before 1987. Therefore, the risk-free rate proxy is based on the same measure for the sample period as a whole, instead of incorporating Norwegian treasuries in the first part of the sample. In doing so, we maintain consistency for the thesis. A time-series plot of the one-month NIBOR can be viewed in Appendix B.

4.3 The Market Proxy: Benchmark Index

The benchmark used in mutual fund performance evaluation is crucial, as previous research has shown; see the literature Section. One needs to establish a benchmark that represents a fitting market return. In an ideal world, the market portfolio from CAPM would be used. However, this is not readily available. Therefore, we need a reasonable approximation of the market portfolio. Oslo Stock Exchange provides several indices that could serve as proxies for the market return. The Oslo Stock Exchange Benchmark Index (OSEBX) consists of the most-traded shares on Oslo Stock Exchange. Oslo Stock Exchange All-Share Index (OSEAX) consists of all shares listed on the exchange. The OSEAX is adjusted for dividend payments and corporate actions daily. In addition, there is a Mutual Fund Index (OSEFX) at Oslo Stock Exchange, frequently used as a benchmark by Norwegian mutual funds in their reports. The OSEFX serves as a natural benchmark, complying with UCITS directives in addition to being designed to meet diversification requirements.

However, the OSEFX originated in 1995, whereas the mutual fund sample origins in 1987. Thus, the OSEFX cannot function as a market proxy in the years from 1987 to 1994. In the investigation of the fund performances, we want to incorporate only one benchmark for the whole period, 1987-2019.

Based on this, the most fitting market proxy is the OSEAX. We thus use the OSEAX as the market proxy. Data for OSEAX is provided by Ødegaards database (Ødegaard, 2020) and described in (Ødegaard, 2021a), where the least liquid and smallest stocks are filtered out.

4.4 Risk Factors

Data on risk factors for the Carhart (1997) four-factor model and the Fama-French three-factor model is provided by Ødegaard based on empirical data from the Oslo Stock Exchange. Ødegaard has provided a database that contains data on all the risk factors needed, including the market factor (MKT), Small-Minus-Big (SMB), High-Minus-Low (HML), and lastly, the one-year momentum (PR1YR) (Ødegaard, 2020).

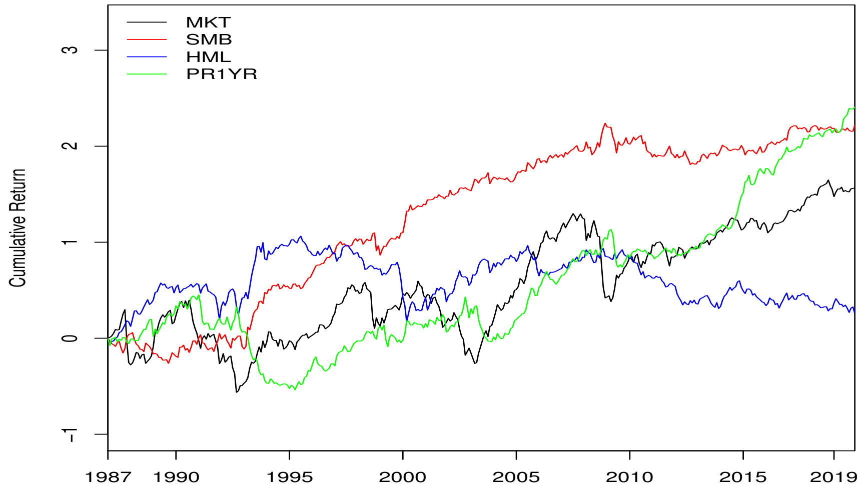
The data in Ødegaards database is described in Ødegaard (2021a) and Ødegaard (2021b). The SMB factor considers small-capitalization companies' average return minus big (large) capitalization companies' returns. HML consists of the difference between the average return on high book-to-market ratio portfolios and the average return of low book-to-market ratios, that is, value minus growth. The PR1YR is the prior-one-year factor, calculated as the difference between the highest one-year lagged return and lowest one-year lagged return in holding the stock. Note that shorting of stocks is not considered for this factor. Factor loadings for each individual fund are provided in Appendix A Table A.II.

In the investigation of the funds' ability to generate returns in excess of the risk factors, it is important to consider the returns of the risk factors themselves. Figure 1 shows the cumulative log-returns of the risk factors from inception to 2019. We see that the momentum factor generated the highest cumulative return in the period, followed closely by the SMB factor. The market excess return factor is volatile in the years leading up to 2010, past this point it moves in a narrower channel upwards. Cumulative log-return of the market proxy, not subtracted the risk-free rate, is shown in Figure 2 panel A. Lastly, the HML factor has the lowest cumulative return out of the risk factors considered.

We now turn to other properties of the risk factors, taking a look at the mean return, standard deviation, maximum monthly return, and minimum monthly return. This is calculated for each risk factor, and we evaluate them through the whole sample period. In addition, we investigate four sub-sample periods. Lastly, we look at the correlation between the risk factor returns for the period as a whole.

We see, as indicated by the cumulative log-returns in Figure 1, that the momentum factor has the highest mean return for the sample period as a whole, followed by the SMB- and market factor, respectively. What is also interesting to note, is that the momentum factor has

Figure 1: The figure presents the cumulative log-returns of the risk factors considered in this thesis. The cumulative log-returns are calculated from inception in 1987 for each month throughout the whole sample period. The MKT factor represents the excess return of the OSEAX (black). The cumulative log-return of SMB and HML is shown in red and blue lines, respectively. Finally, the momentum (PR1YR) factor is presented in green.



the second narrowest range of returns, with a max of 15.43% and a min of -16.78%, second to the HML factor. The market factor has the highest standard deviation among the risk factors. Descriptive statistics of the market proxy not subtracting the risk-free rate is shown as OSEAX in Table 2. The SMB factor has the lowest standard deviation for the period as a whole. Still, the SMB factor has the second-highest mean return.

In the first sub-period, 1987-1995 Q1, the HML factor generated a mean return of 14.15% after this sub-period. Although the mean return of the factor was negative for the following three periods. The standard deviation of the HML factor decreases throughout the four sub-periods as well as the range of its monthly returns. Looking at the market factor, we see that this has generated a mean return above 10% in the two sub-periods following 2003 Q3. For the two preceding sub-periods, however, the mean return was 1.89% and 3.20%. We see that the SMB factor generates the highest sub-period mean returns in the first and second sub periods. Also, the mean return of the factor does not fall below zero for any of the sub-periods.

Results from the correlation matrix in Table 1 tell us that the correlation between the risk factors is close to zero, except for the correlation between the market factor and SMB factor. Here, we see a negative correlation of -0.45, telling us that when the market excess return is positive, the SMB factor return tends to be negative, and vice versa.

Table 1: This table presents statistics of the risk factors for five different periods: 1987-2019, 1987 Q1-1995 Q1, 1995 Q2 - 2003 Q2, 2003 Q3 - 2011 Q3, and 2011 Q4 - 2019 Q4. Columns 2 and 3 contain the annualized mean return and standard deviation for each factor, respectively. The maximum monthly return in a given period is shown in column 4, whereas column 5 shows the lowest return in a given period. All numbers are given in percent. Lastly, in panel F, a correlation matrix is given for the whole sample period (1987-2019).

	Mean Return	Standard Dev.	Max	Min
Panel A: 1987 - 2019				
MKT	7.09	20.74	16.51	-28.69
SMB	7.85	14.16	22.14	-17.08
HML	2.08	16.18	14.66	-16.65
PR1YR	8.67	16.31	15.43	-16.78
Panel B: 1987 Q1 - 1995 Q1				
MKT	1.89	25.17	16.51	-28.69
SMB	8.36	17.53	22.14	-10.46
HML	14.15	19.55	14.66	-15.72
PR1YR	-4.83	18.14	13.50	-16.78
Panel C: 1995 Q2 - 2003 Q2				
MKT	3.20	20.24	11.81	-23.27
SMB	14.27	12.92	13.27	-17.08
HML	-0.63	18.09	9.75	-16.65
PR1YR	9.23	18.54	15.43	-14.22
Panel D: 2003 Q3 - 2011 Q3				
MKT	12.37	23.53	14.88	-24.58
SMB	5.14	14.94	12.82	-11.03
HML	-2.07	13.95	9.10	-12.24
PR1YR	11.27	14.84	9.84	-16.09
Panel E: 2011 Q4 - 2019 Q4				
MKT	10.91	11.37	9.96	-8.26
SMB	3.64	10.26	10.33	-7.42
HML	-3.13	11.57	6.40	-7.42
PR1YR	19.03	12.38	12.05	-8.82
Panel F: Correlation Matrix 1987 - 2019				
	MKT	SMB	HML	PR1YR
MKT	1			
SMB	-0.45	1		
HML	0.05	-0.14	1	
PR1YR	-0.16	0.11	-0.11	1

4.5 Mutual Fund Returns: Potential Biases

Survivorship bias is one of the most obvious biases, which occurs if one omits funds that close during the period of investigation (Brown, Goetzmann, Ibbotson, & Ross, 1992). Specifically, the exclusion of terminated funds can lead to survivorship bias through sample selection. Closed funds have implemented strategies that have been proven to fail. In exclusion of these closed funds, the approach would also be eliminated (Elton et al., 1996). Funds are not primarily closed due to a poor one-year performance. More often, funds close due to multiple years of underperformance (Carpenter & Lynch, 1999). If one included only the surviving funds, one could see the sample average return being upward biased, due to the fact that funds yielding a high return and persistently performing tend to be survivors. As the aggregate performance becomes artificially inflated, it leaves the sample results incomplete.

Carpenter and Lynch (1999) also introduces a look-ahead bias, and they address the importance of year-end returns. Look-ahead bias can occur if funds are required to exist over a given period, which would exclude funds with too short lifetimes. Funds are not necessarily closed due to poor performance; one could argue that funds are closed due to the cost of operating the fund surpassing the profits. However, this may come as a result of poor performance of the fund. In order to prevent survivorship bias, both closed funds and funds still open are included in the data set. The funds are required to have at least 12 observations to be included, thus we might experience some look-ahead bias.

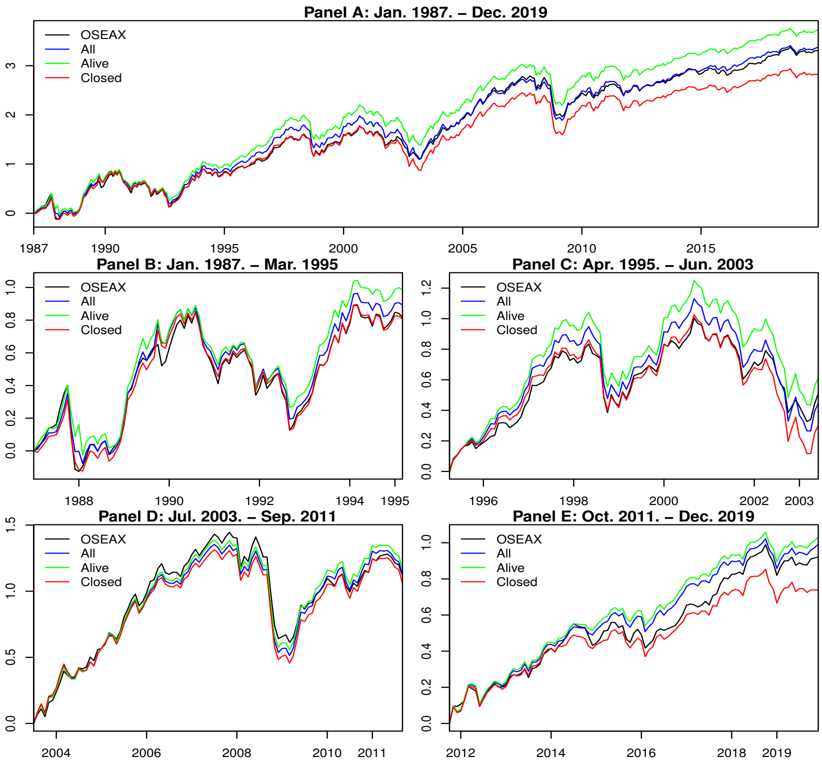
Note that Harvey and Liu (2020) argues that short return series may cause outliers in the t-statistics of alpha through the bootstrapping procedure, which reduces the power of the test. Therefore, the requirement of a minimum number of observations is a double-edged sword, where look-ahead bias may occur with requirements of too many observations, but lowering the requirements may reduce the power of the test. Fund size is not regarded in the restrictions, all sizes are eligible for inclusion. In the case of merging funds, money is assumed to be invested in the acquiring fund (Elton et al., 1996).

We look at the cumulative log-returns of the OSEAX, and equally weighted portfolios consisting of: all funds available in a given month, all funds available in a given month that are still open as of 31.12.2019, and all funds available in a given month that were closed prior to 31.12.2019. We see for the period as a whole, that the equally weighted portfolio of funds still open as of 31.12.2019 has the highest cumulative log-return, presented in panel A of Figure 2. Throughout the sub-periods, we also see that this portfolio generates the highest cumulative return. However, we note that it was lower than the OSEAX in the years 2005-2009, as seen in panel D Figure 2.

Looking at the equally weighted portfolio of all funds shown in Figure 2, we see that it lies close to the cumulative log-return of the OSEAX. From the sub-periods plotted in panel B-E, we see that the all-funds portfolio oscillates around the OSEAX cumulative log-return, but comes out marginally better at the end of the periods.

Finally, we look at the portfolio of funds that have been closed in Figure 2. We see that this portfolio consistently generated a cumulative log-return below the other three return series investigated. This finding tells us that the funds that have closed throughout the period might have a downward effect on the aggregate return of funds to such a degree that the outperforming funds are hidden by the bad returns of the funds closed.

Figure 2: The figure shows cumulative log-returns for OSEAX and three equally weighted portfolios of funds: containing all funds, the funds alive as of 31.12.2019, and the funds who have been closed throughout the whole sample period. Five periods are presented: the sample period as a whole, (1987-2019), and four sub-periods of 8 years and a quarter spanning, 1987 Q1-1995 Q1, 1995 Q2 - 2003 Q2, 2003 Q3 - 2011 Q3, and 2011 Q4 - 2019 Q4. Presented in Panels A-E respectively.



We also look at some descriptive statistics of the portfolios presented in Figure 2. Table 2 contains the mean return, standard deviation, skew, kurtosis, and the maximum and minimum monthly return in a period. From the cumulative log-returns presented in Figure 2 we noted that the portfolio of funds open as of 31.12.2019 had the highest cumulative return. The same holds for the mean return of the portfolio for the given periods.

The standard deviation of the open funds portfolio is relatively high. As for higher moments, we see that we have some negative skew and excess kurtosis in relation to the normal distribution. As noted in the discussion of the cumulative log-returns in Figure 2, we see that the portfolio of all funds and the OSEAX generated mean returns and standard deviations that were close throughout the sub-periods. The standard deviation of the OSEAX and the portfolio of all funds are close to each other for the periods. Both exhibit negative skew and excess kurtosis in relation to the normal distribution.

Finally, the portfolio of funds that have been closed generated the lowest mean return across all panels. In addition to having the highest standard deviation, we here, as in the other cases, see negative skew and excess kurtosis.

Table 2: The table displays descriptive statistics of the OSEAX (the benchmark used in this thesis) and for equally weighted portfolios of funds consisting of all funds available in a given period, all funds available in a given period that is still open as of 31.12.2019, and lastly, all funds available in a given period that closed prior to 31.12.2019. Five panels are presented, one for each time period under investigation. That is, the sample period as a whole, 1987-2019, and four sub-periods, 1987 Q1-1995 Q1, 1995 Q2 - 2003 Q2, 2003 Q3 - 2011 Q3, and 2011 Q4 - 2019 Q4, in panel A-E, respectively. Four moments of the return series are presented, mean return, standard deviation, skewness, and kurtosis in columns 2-5, respectively. In addition, the maximum and minimum monthly returns are presented in columns 7 and 8, respectively. Mean return and standard deviation are annualized and in percent, whereas max- and min-returns are percentages.

	Mean Return	Std.Dev.	Skew	Kurtosis	Max	Min
Panel A: 1987 - 2019						
OSEAX	12.38	20.62	5.84	-0.99	17.45	-27.42
All	12.60	20.63	5.11	-0.81	17.55	-25.28
Alive	13.71	20.81	5.04	-0.79	18.65	-25.39
Closed	10.83	20.89	5.15	-0.79	18.14	-25.09
Panel B: 1987 Q1 - 1995 Q1						
OSEAX	12.63	3.65	4.71	-0.95	17.45	-27.42
All	13.03	3.76	3.39	-0.47	17.55	-19.56
Alive	14.38	4.15	3.27	-0.50	18.65	-17.03
Closed	12.02	3.47	3.74	-0.49	18.14	-21.48
Panel C: 1995 Q2 - 2003 Q2						
OSEAX	9.09	2.63	4.87	-0.75	12.49	-22.57
All	8.87	2.56	4.28	-0.72	13.82	-22.83
Alive	10.96	3.17	4.26	-0.64	16.11	-23.12
Closed	7.08	2.04	4.31	-0.75	14.10	-22.64
Panel D: 2003 Q3 - 2011 Q3						
OSEAX	15.53	4.48	5.27	-1.08	15.05	-23.93
All	15.33	4.43	5.26	-1.07	15.54	-25.28
Alive	15.87	4.58	5.31	-1.08	15.62	-25.39
Closed	14.58	4.21	5.20	-1.06	15.42	-25.09
Panel E: 2011 Q4 - 2019 Q4						
OSEAX	12.26	3.54	3.74	-0.40	10.19	-8.10
All	13.16	3.80	4.51	-0.59	10.10	-9.36
Alive	13.62	3.93	4.59	-0.58	10.31	-9.34
Closed	9.65	2.79	4.15	-0.45	9.71	-9.39

5 Empirical Results

This Section presents findings of the empirical analysis of Norwegian mutual fund performance. First, aggregate and individual performance are investigated through the use of factor models. Three factor models are presented: single-factor as in equation (1), the three-factor model from equation (4), and the four-factor model as in equation (5).

The main focus is on the four-factor model, as this is applied in all bootstrapping procedures, the other models are used for comparison. Next, bootstrap results in the Kosowski et al. (2006) and Fama and French (2010) framework are examined to determine if we have evidence of skilled or unskilled managers in our sample. Finally, we turn to the double-bootstrap results, investigating whether funds are able to outperform the market net-of-fees.

5.1 Factor Model Fund Performance on Aggregate and Individual Levels

As a means of measuring the aggregate performance of the Norwegian mutual funds, an equally weighted portfolio of the funds is created. That is, all funds alive in a given period are put into a portfolio where each fund is given the same weight. Under the assumption of true alpha being zero, the focus is on whether funds on aggregate have been able to yield a positive alpha.

Ordinary Least Squares (OLS) regressions have been run in order to estimate factor coefficients and the abnormal performance of the equally weighted portfolio of funds. Table 3 shows estimated results of regression analyses using the three factor models presented in this thesis. This is conducted for the sample period as a whole, and four sub-periods of equal length: 8 years and one quarter.

Findings from the regression analysis tell us that the market factor has the highest impact on the funds' return, and a highly significant one at that. For all models and all time periods where the models are applied, we see a statistically significant market coefficient. Adjusted R-squared from the models tells us that the risk factors applied in the models explain between 88-96 percent of the variation in the aggregate fund returns.

Table 3: This table shows estimated alpha, annualized and in percent, factor loadings, and adjusted R-squared for an equally weighted portfolio of funds. All funds open in a given period are included in the equally weighted portfolio. Single-factor-, three-factor-, and four-factor- estimates are provided in rows 1-3 of each panel, respectively. Panel A provides full sample estimates whereas panel B-E provides sub-sample estimates. Standard errors are shown in parentheses below the estimate. Stars, *, **, ***, represent statistical significance of 1%, 5%, and 10% respectively. Alpha and factor loadings are presented in columns 2-5, whereas column 6 provides the adjusted R-squared of the model.

Model	α	β_{MKT}	β_{SMB}	β_{HML}	β_{PR1YR}	$\bar{R}^2$
Panel A: 1987-2019						
Jensen's	0.5	0.96***				0.92
Single-factor	(0.49)	(67.10)				
Fama-French	-0.38	0.99***	0.10***	-0.06***		0.93
Three-factor	(-0.37)	(64.54)	(4.34)	(-3.39)		
Carhart	0.04	0.99***	0.10***	-0.07***	-0.04*	0.93
Four-factor	(0.03)	(64.13)	(4.43)	(-3.66)	(-2.48)	
Panel B: 1987-1995 first quarter						
Jensen's	0.69	0.87***				0.89
Single-factor	(0.26)	(28.28)				
Fama-French	-1.20	0.89***	0.11*	0.06		0.90
Three-factor	(-0.44)	(24.50)	(2.36)	(1.48)		
Carhart	-1.20	0.89***	0.11*	0.06	0.00	0.90
four-factor	(-0.44)	(24.29)	(2.35)	(1.46)	(0.04)	
Panel C: 1995, 2nd Q-2003, 2nd Q						
Jensen's	-0.45	1.07***				0.94
Single-factor	(-0.24)	(38.52)				
Fama-French	-2.62	1.07***	0.15***	-0.08*		0.95
Three-factor	(-1.44)	(40.08)	(3.51)	(-2.50)		
Carhart	-2.1	1.05***	0.17***	-0.08**	-0.08**	0.95
four-factor	(-1.18)	(38.88)	(4.02)	(-2.89)	(-2.75)	
Panel D: 2003, 3rd Q - 2011, 3rd Q						
Jensen's	-0.15	1.0***				0.96
Single-factor	(-0.08)	(45.76)				
Fama-French	-0.65	1.01***	0.04	-0.07*		0.96
Three-factor	(-0.36)	(36.14)	(0.97)	(-1.87)		
Carhart	0.13	1.03***	0.08.	-0.08*	-0.10**	0.96
four-factor	(0.07)	(37.6)	(1.81)	(-2.16)	(-3.01)	
Panel E: 2011, 4th Q - 2019						
Jensen's	2.04	0.89***				0.87
Single-factor	(1.46)	(26.04)				
Fama-French	1.46	0.91***	0.04	-0.09**		0.88
Three-factor	(1.06)	(25.01)	(0.93)	(-2.87)		
Carhart	0.70	0.92***	0.05	-0.09**	0.03	0.88
four-factor	(0.44)	(23.64)	(1.13)	(-2.87)	(0.94)	

From the latest sub-period, 2011 Q4-2019 Q4, we see the highest annualized alpha estimates, in addition to the lowest adjusted R-square in comparison to the other sub-periods. The SMB factor, which is included in both the Fama-French three-factor model and Charharts four-factor model, is statistically significant for the period as a whole and the first two sub-periods. This tells us that the aggregate fund portfolio is impacted to a higher degree in the first part of the sample period. The HML factor tells a slightly different story, being insignificant in only the first sub-period, 1987-1995 Q1. Lastly, the momentum factor, PR1YR, included only in the four-factor model, is significant for the sample period as a whole. The PR1YR factor is non-significant only in two sub-periods.

To investigate the abnormal performance on the aggregate level, a plot of the evolvement of alpha for the equally weighted fund portfolio is provided in Figure 3. Here, both extending and rolling 36-month alpha is plotted, for both the Jensen's alpha model and Charharts four-factor model. In addition to the alpha estimate, standard error bands and Newey-West standard error bands are plotted.

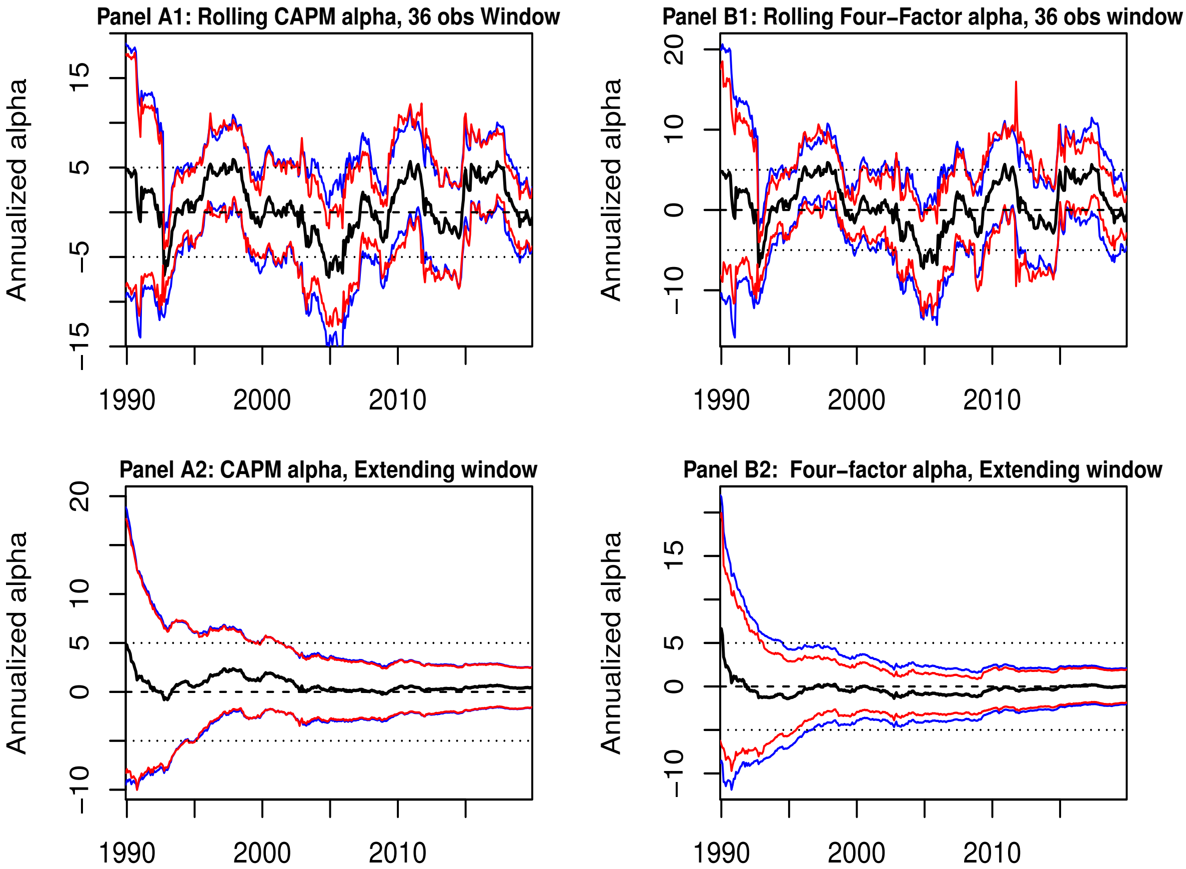
From the rolling Jensen's alpha, panel A1, we see that the annualized alpha moves between 5% and -5%. The standard error bands move closer to the estimate itself throughout the sample period, telling us that the variation in alpha is lower. In addition, the channel of the standard error bands tells us about the trend of the alpha. If the channel is narrowing and pointing upwards to the right, the alpha is in a strong trend upwards.

In the extending-window estimation of alpha starting at 36 months, however, we see that the abnormal performance is barely below zero in the early '90s. We also note that from early 2000 up to 2019, the extending-alpha experiences less volatility, this can be seen by the standard error bands narrowing, and by the decrease in movement in the estimate itself.

Inspection of the 36-month rolling-window of abnormal performance using the four-factor model reveals that the abnormal performance is less volatile than the corresponding alpha of the Jensen's alpha model. A visual inspection of the rolling-alpha indicates that the estimate moves in the interval of -5% to 5% throughout the period. Yet we find more observations at a positive rolling alpha of 5% than -5%.

The extending-window of abnormal performance from the Carhart four-factor model tells us a similar story to that of the Jensen's alpha. The abnormal performance estimate volatility reduces as more observations are put into the estimation sample. However, in application of the four-factor model, we see that the abnormal performance falls below 0 more frequently when compared to the Jensen's extending-window alpha.

Figure 3: The figure displays annualized 36 observation rolling-window alpha and extending-window alpha, starting at 36 months. Alpha is given in percent (black). The alpha is estimated for an equally weighted portfolio of the funds operating in a given month. Jensen's alpha model is applied in panel A, whereas panel B alpha is estimated through the use of the Charhart four-factor model. Standard error bands (blue) and Newey-West standard error bands (red) are added to the plot for visualization of the variation and trend in alpha.



In Table 4 the top and bottom individual performers ranked on t-statistic of alpha is presented. We see that individual funds are able to generate abnormal returns beyond covering their fees for the investors. The top 7 performers generate statistically significant positive abnormal returns. In addition, we note that the top 8th and 9th performers generates a p-value of alpha below 0.1.

As for the bottom-performing funds, all generate statistically significant negative returns net of fees at the 5% significance level. We also see that the bottom four funds generate a p-value of 0.01 or below.

Table 4: The table shows individual parametric estimates for the top 10 and bottom 10 funds ranked on the t-statistic of alpha. Panel A presents the Top performers, along with the corresponding alpha and p-value of alpha in the second and third row of the panel respectively. Panel B shows the individual estimates of the bottom 10 performers t-statistic of alpha of the fund in the first row. The following two rows present the estimated alpha and p-value corresponding to the t-statistic respectively.

Panel A: Top performers, ranked on t-statistic of alpha										
	Top	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
t-statistic	2.23	2.19	2.15	1.94	1.94	1.83	1.82	1.43	1.36	1.24
Alpha	3.02	10.95	5.38	3.78	3.26	1.95	7.62	6.18	1.60	4.18
p-value	0.01	0.01	0.02	0.03	0.03	0.03	0.03	0.08	0.09	0.11
Panel B: Bottom performers, ranked on t-statistic of alpha										
	Bottom	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
t-statistic	-3.22	-3.18	-2.96	-2.63	-2.45	-2.14	-2.03	-1.99	-1.89	-1.81
Alpha	-12.25	-16.93	-5.31	-6.44	-4.28	-4.98	-3.32	-8.54	-19.06	-18.62
p-value	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.02	0.03	0.04

Summarizing the results of the aggregate and individual fund performance findings suggest that on aggregate, mutual funds are not able to generate statistically significant alpha different from zero. As we are investigating the net-of-fees returns, this tells us that on aggregate funds do not generate high enough returns to cover their costs on aggregate.

The individual fund performances measured through standard parametric OLS regression tell us that the top performers are able to generate positive abnormal returns. Findings from the bottom-performing funds tell us that these funds indeed generate returns significantly below zero. From this, we gather that the top funds generate alpha large enough to add value for investors after fees have been deducted. Bottom performers are, however, not able to generate returns covering the cost of investing in these funds.

5.2 Kosowski et al. (2006) Approach, Luck or Skill?

We now turn our investigation into whether the performances observed are caused by luck or skill. Kosowski et al. (2006) apply a baseline bootstrap where residuals are assumed to be independently and identically distributed. In this thesis, however, the before-mentioned argument of Politis and Romano (1994) is followed. Therefore, we use a block length of 4 in the main bootstrapping procedure. That is, cross-sectional dependence in residuals is taken into consideration.

Table 5 displays bootstrapped p-values of the Kosowski et al. (2006) method, using a block length of four in the bootstrapping procedure. Panel A shows funds ranked on their alpha estimate, whereas in Panel B, funds are ranked based on their t-statistic of alpha. Parametric, single-test p-values tell us that the top- and bottom performers are significantly outperforming and underperforming the market. However, taking a look at the top performers ranked on alpha, we see that these do not show any ability of stock-picking skill with a bootstrapped p-value of 0.52.

According to the bootstrapped p-values of all the top performers inspected here, we see no evidence of stock-picking skill. However, when looking at the bottom performers ranked on alpha, bootstrapped p-values tell a different story. When the left tail performance is inspected, we see evidence of a lack of skill in stock-picking. All five left tail performers considered in Panel A provide significant evidence, with a bootstrapped p-value beneath 5% for all percentiles presented.

Inspecting Panel B, where funds are ranked based on their t-statistic of alpha. We see that, here as well, top funds fail to provide evidence of stock-picking skill, as opposed to luck. With a bootstrapped p-value as high as 0.8 for the top-performing fund when ranked on t-statistic of alpha, we cannot say that this performance is due to anything but luck.

When we however inspect the left tail performers when ranked on the t-statistic of alpha, we see clear evidence of a lack of stock-picking skill among the worst Norwegian mutual fund managers. All bottom performers considered here display significant evidence of underperformance due to lack of skill bar the absolute bottom performer.

Table 5: This table provides cross-sectional bootstrap results from the Kosowski et al. (2006) block bootstrap method for all the Norwegian mutual funds in the sample 1987-2019. Both panels provide findings for various percentiles of both top- and bottom performances. Panel A ranks funds based on four-factor alpha and provides four-factor alpha estimates, along with bootstrapped p-values and parametric p-values of the corresponding percentiles. Panel B shows and ranks performance on the t-statistic of alpha, along with the bootstrapped p-value and parametric p-value. Columns 2-6 report statistics of top funds, while columns 7-11 contain the bottom funds. The first rows in the two panels reports estimated alpha and t-statistic of alpha, respectively. Row 2 and 3 contain bootstrapped p-values and parametric p-values of the funds in both panels. The statistics are based on 10,000 bootstrap resamples and ranked on their parametric t-statistic of alpha in both panels.

	Top	2.nd	3.rd	Top 5%	Top 10%	Bottom 10%	Bottom 5%	3.rd	2.nd	Bottom
Panel A: Fund Ranked on Four-Factor Model Alpha										
Alpha	10.95	7.62	6.18	5.38	3.02	-7.01	-11.89	-16.93	-18.62	-19.06
Bootstrapped p-value	0.52	0.08	0.57	0.29	0.80	0.01	0.00	0.02	0.01	0.01
Parametric p-value	0.01	0.03	0.08	0.02	0.01	0.08	0.05	0.00	0.04	0.03
Panel B: Fund Ranked on t-statistic of Four-Factor Model Alpha										
t-statistic	2.23	2.19	2.15	1.94	1.19	-1.73	-2.45	-2.96	-3.18	-3.22
Bootstrapped p-value	0.80	0.52	0.29	0.17	0.70	0.01	0.00	0.00	0.02	0.19
Parametric p-value	0.01	0.01	0.02	0.03	0.12	0.04	0.01	0.00	0.00	0.00

We would also like to inspect the cross-sectional distribution of the t-statistics of alpha, in order to elaborate on the findings in Table 5. Figure 4 shows bootstrapped t-statistics of alpha distributions for various percentiles. We compare the 10.000 bootstrapped t-statistics at given percentiles in the top and bottom of performances, to the actual estimated t-statistic for that same percentile in the actual t-statistic of alpha distribution. In addition to this, we also present what the t-statistic of alpha had to be in order to generate a bootstrapped p-value of 5%, all else equal.

In Panel A, we see the top performer, fifth percentile, and tenth percentile, whereas Panel B shows the left tail performance, e.g. bottom performer, bottom fifth percentile, and bottom tenth percentile. For the top performer ranked by the t-statistic of alpha, presented in panel A1, we see that the actual t-statistic lies to the left of the peak of the distribution of bootstrapped t-statistic for that same percentile.

The actual t-statistic of alpha, and what it had to be in order to generate a bootstrapped p-value of 5% all else equal, differs significantly from each other with an absolute difference of 1.47 (3.7 - 2.23). We note, as seen from the bootstrapped p-values from Table 5, that there are a considerable amount of bootstrapped t-statistics generated that are larger than the actual t-statistic of alpha for the top fund.

We also see that the bootstrapped t-statistics are left-skewed, and range from a little above

1 to above 5 in the tails. This evidence further indicates that the returns generated by the top-performing fund is due to luck rather than stock-picking skill. In panels A2 and A3, we see less evidence of skew, and in addition a smaller range of the bootstrapped t-statistics. In panel A2, the span ranges from just under 1 to approximately 3. Panel A3 displays a range between approximately zero to just above 2. Yet, we find a considerable amount of bootstrapped t-statistics above the actual t-statistic of alpha, confirming the results from the bootstrapped p-values of these funds.

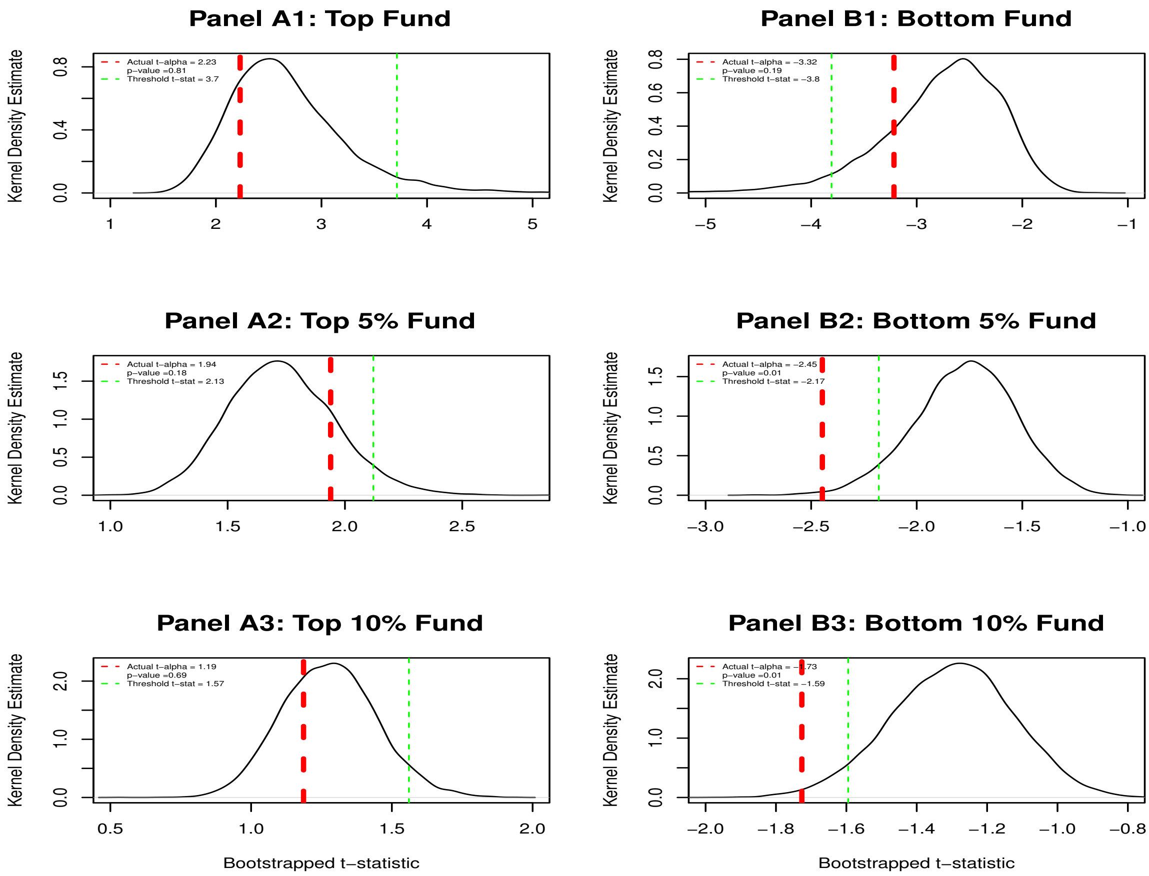
Panel B shows the left tail of funds ranked on the t-statistic of alpha. For the bottom fund (Panel B1), we see a right skew in the density of the bootstrapped t-statistics, and a range from above -5 to just under -1. The actual t-statistic of the bottom performer lies to the left of the peak of the bootstrapped t-statistics of the bottom performer. However, we see that the bootstrap procedure produces more t-statistics below the actual t-statistic of alpha than what would warrant a rejection of the null hypothesis for the bottom performer.

From the other two panels, B2 and B3, we see that the actual t-statistics are located far out in the tail of the bootstrapped t-statistics. The bootstrapped t-statistics here, as in panel A2 and A3, seem less skewed than the bottom percentile distribution of bootstrapped t-statistics. With a narrower range and less extreme observations in the tails, we see that there is a very small amount of bootstrapped t-statistics below the actual t-statistics at these percentiles.

In Figure 5, two panels are presented: The probability density function of both actual and bootstrapped t-statistics in Panel A, and Panel B presents the cumulative distribution function of the actual and bootstrapped t-statistics of alpha. When comparing the bootstrap generated cross-sectional distribution of t-statistics of alpha to the actual cross-section of t-statistics of alpha, we especially note the differences in the shape of the two distributions.

We see that the actual t-statistics of alpha exhibit a lower peak and fatter tails, especially the left, and are more skewed to the left than the bootstrapped t-statistics. In addition, we see a complicated feature in the distribution of the t-statistics of alpha: Two "shoulders" around zero. That is, we do not find the classical bell-shaped distribution. In an ideal world, where we can reject the null of no abnormal performance among Norwegian mutual fund managers, the distribution of actual t-statistics of alpha would lie to the right of the distribution of bootstrapped t-statistics of alpha. This is, however, not the case here. Where actual t-statistics of alpha is to the left of the bootstrapped t-statistic distribution. Indicating luck,

Figure 4: The panels display various percentiles in the cross-section of the t-statistics of alpha. The solid lines are the kernel densities of the bootstrapped four-factor t-statistics of alpha. The actual estimated fund alpha is shown as the vertical dotted line. Panels A1-A3 present top performance distributions for the top fund, top 5 percentile and top 10 percentile respectively. Panels B1-B3 report left-tail performance for the bottom, bottom 5 percentile, and bottom 10 percentlie funds, respectively. Statistics are based on 10,000 bootstrap resamples, and are ranked on the t-statistic of alpha in both panels.



rather than skill for top performing funds, and a lack of skill in the left tail.

The bootstrap offers the ability to deal with complex shapes of the full cross-sectional distribution of t-statistics. In addition, we are able to get more accurate measures on fat vs. thin tails in the actual distribution through the use of bootstrapping. Finally, this results in differences between the inferences based on normality assumptions and the bootstrap procedure, as a result of the lack of normality in the cross-sectional distribution of the actual t-statistics of alpha for the funds.

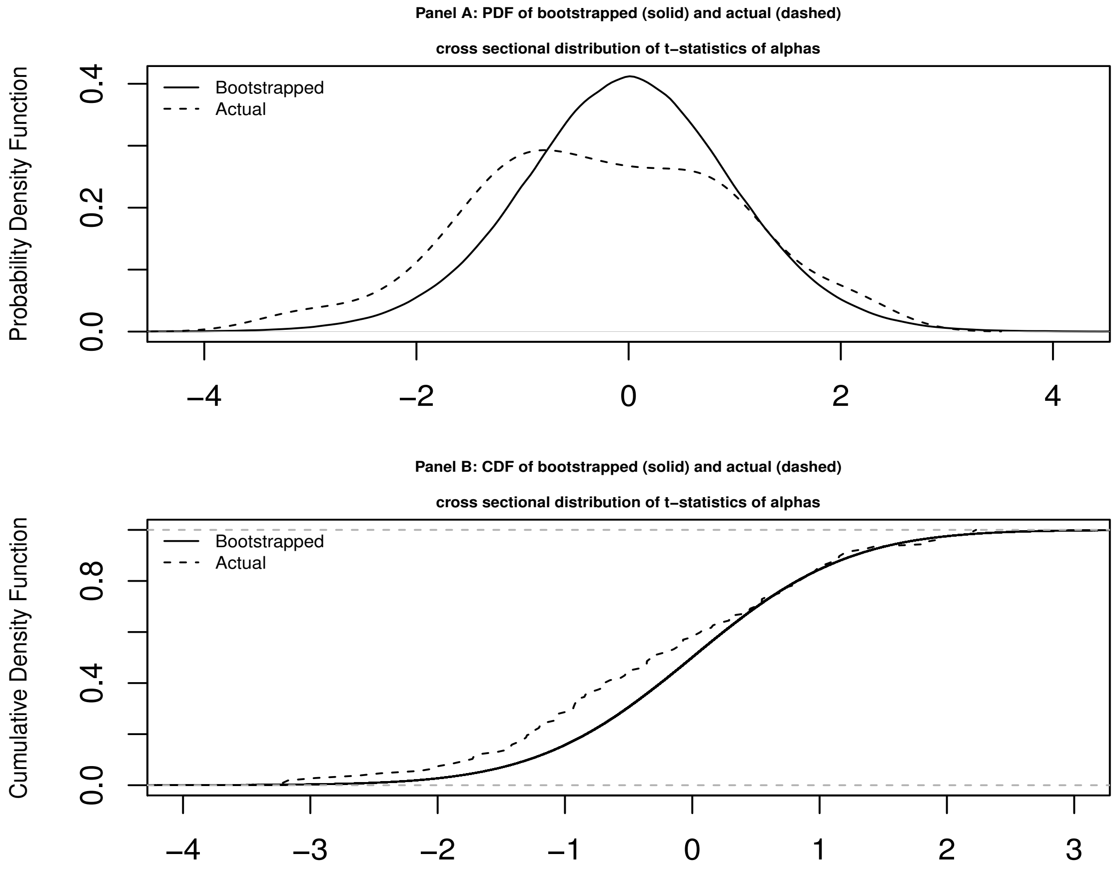
We further elaborate on the results in Panel A with the cumulative probability function of actual and bootstrapped t-statistics of alpha presented in Panel B. Panel B show us that there is more probability mass in the left tail of the actual distribution in comparison to the bootstrapped t-statistic distribution. We see that the cumulative distribution of actual t-statistic lies above the bootstrapped t-statistics from about -3.2 up to 0.2. In the right tail, we see that the actual t-statistic distribution dips below the bootstrapped distribution, but for the main part it clings tightly to the bootstrapped cumulative distribution of t-statistics.

The distributions in Figure 5 tells us that the actual t-statistics of alpha are to the left of the bootstrapped t-statistic distributions, which indicates an underperformance in relation to the bootstrap. The evidence is also in line with prior findings from Table 5.

From parametric evidence, we have supporting findings of outperformance. However, when conducting the bootstrap procedure as proposed by Kosowski et al. (2006), findings suggest that superior performance is due to luck rather than stock-picking skill. Further, we have found bootstrapped p-values for the bottom performers that are so low that we can support the rejection of the null hypothesis. That is, the bootstrap evidence from this section tells us that the underperformance is not due to bad luck alone, but suggests that underperformers inhabit a lack of stock-picking skill.

From this analysis, we infer that the top-performing Norwegian mutual fund managers are not skillful enough to generate abnormal returns beyond covering their fees. As for the Bottom performers, we see that they in fact lack stock-picking skills. Meaning that these managers are not skillful enough to cover their fees.

Figure 5: The figure shows two panels illustrating the probability density function (PDF) and cumulative density function (CDF) of the estimated and bootstrapped t-statistics. Panel A displays the PDF of bootstrapped- (solid) and actual (dotted) t-statistic of alpha. Panel B shows the CDF of the actual (dotted) and bootstrapped t-statistics of alpha (solid). Full sample four-factor estimates are used for the actual t-statistics of alpha.



5.3 Sensitivity Analysis Kosowski et al. (2006)

In the methodology of Kosowski et al. (2006), they apply multiple bootstrap procedures in order to evaluate the robustness of their findings. In this sub-section, some of the robustness tests conducted in the paper of Kosowski et al. (2006) are applied. When the bootstrapping procedure is changed, specific assumptions may be affected and the null hypothesis may be changed. Such changes and their effects will be addressed in the corresponding sections below.

5.3.1 Time-series dependencies

Politis and Romano (1994) argued that dependence in return residuals should be allowed through adopting the stationary bootstrap method. The main bootstrap approach conducted in this thesis keeps the dependence in return residuals. However, we here apply different lengths of block-resampling.

When applying the stationary bootstrap method, it is essential to determine the block length. The block length is determined through the asymptotic formula: $l\tilde{T}^{\frac{1}{h}}$, where T is the number of observations, and h equals 3, 4, or 5. As argued by Hall, Horowitz, and Jing (1995). In their study, they found that h depends on the context. In the context of one-sided distribution functions for the given test statistic, the block bootstrap estimator is given as h=4. For a two-sided distribution, Hall et al. (1995) argued that $h = 5$ and $h = 3$ should be chosen when determining block bootstrap estimator variance. In the situation of this thesis, this generated a block length of 4.

When applying this approach, blocks of random length are resampled and draws a string of independent and identical random variables from a geometric distribution. This process arranges the blocks to output a stationary pseudo-time-series. Results are compared under different block lengths of 1, 4 (the block length used in the main investigation of the thesis), and two random block lengths (2 and 10).

Results of the different block length bootstraps are presented in Appendix D. From the results of the different block length procedures, we find minimal differences. Therefore, we infer that the results are robust.

5.3.2 Portfolios of Funds

Kosowski et al. (2006) argue that in order to determine whether individual cross-sectional fund alpha analysis is affecting inference, one has to consider the corresponding average statistics in each tail for portfolios of funds. This is conducted for the bootstrap procedure applied in Section 5.2.

The test conducted here functions as a robustness check for the main results of the thesis. In the application here, the null hypothesis is altered in order to account for portfolios, as opposed to individual funds. The altered null hypothesis state: no abnormal returns (zero alpha performance) across portfolios.

Funds are ranked individually based on the estimated t-statistic of alpha from the four-factor model, before being bundled into portfolios consisting of 2, 3, and 5 funds. Portfolios are bundled from the top, i.e., the top-performing portfolio consist of the two funds with the highest estimated t-statistic of alpha, then the second top-performing portfolio consists of the top 3rd and 4th funds when creating the portfolios of two funds. Alpha and t-statistic of alpha, with both parametric and bootstrapped p-values, are presented in Appendix D, Table D.III.

5.3.3 Length of Data Records

As Harvey and Liu (2020) argue in their evaluation of the Fama and French (2010) error rates, when too short time series are applied, this may significantly alter the results of bootstrapping procedures. They argue that the short time-period requirements in Fama and French (2010) were the main reason for the difference in results between that study and the results in Kosowski et al. (2006). Kosowski et al. (2006) also argued for the use of longer series of returns, in that they identified that short-lived funds tend to have a higher variation and volatility in alpha estimates than that of long-lived funds.

In order to account for the fact that the length of series in the data set may affect the results, requirements of a minimum of 24, 36, and 60 observations have been imposed. Using these new panels of returns, the bootstrapping procedure is conducted once again. Note that when we alter the length of the data records, the number of funds included in the sample is reduced. Specifically, the number of funds in the minimum 24 observation sample is 98, whereas for the minimum 36 and 60 observations samples, we have 96 and 82 funds included, respectively.

Results are reported in Table D.II, Appendix D, where we find minimal differences in the results. Bottom performers provide statistically significant evidence, whereas we find no such evidence for the top performers.

5.3.4 Bootstrap tests for Sub-Periods

In previous sections we have run the bootstrap approach for the time periods as a whole. This section aims to evaluate whether results change when sub-periods are investigated. The funds are still required to have a minimum of 12 observations in the given period in order to be considered for the sample. The entire sample is divided into 4 sub-periods, each consisting of 8 years and one quarter. 1987- Mar. 1995, Apr. 1995 - Jun. 2003, Jul. 2003 - Sep. 2011, finally, Oct. 2011 - Dec. 2019.

Results for each individual sub-period are reported in Table D.IV found in Appendix D, divided into five panels. Panel A represent the full sample period, and panel B, C, D, and E represent the four sub-periods, respectively.

The number of funds in each period varies from the full sample. In the first period, we have 19 funds in the sub-sample, whereas period 2 and 3 have 77 and 68 funds included, respectively. Lastly, for the fourth sub-period we have 76 funds included. The findings for top performers is in line with the results found in Table 5, where we do not reject the null hypothesis for these funds. Also, the performances from the bottom funds in the sub-periods reject the null hypothesis of no true performance. That is, these funds do not have sufficient skill in order to cover costs, and cannot attribute the negative abnormal return to bad luck alone.

5.4 Fama & French (2010) Approach, Luck or Skill?

Bootstrap results from the Fama and French (2010) approach are presented in Table 6, in addition to selected parametric estimates. The table is split into two panels, ranking the funds on estimated alpha from the four-factor model in Panel A, and t-statistic of alpha in Panel B. The key feature in these two panels is the difference between bootstrapped p-values and parametric p-values.

The main difference between Table 6, presented here, and Table 5, from the Kosowski et al. (2006) approach, is the bootstrapping procedure used to generate the bootstrapped t-statistics.

The panels show statistics for various percentiles of the left and right tails in the distributions of alpha and t-statistic of alpha. Specifically, the top-performing fund is followed by the second and third best, in addition to the top fifth and tenth percentile. Preceding the worst performing fund are the tenth and fifth bottom percentiles, along with the second and third bottom performers.

In panel A, the first row shows the estimated alpha, and the second and third rows contain bootstrapped and parametric p-values, respectively. Panel B shows estimated t-statistics of alpha in the first row, followed by (as in panel A), the corresponding bootstrapped p-values and parametric p-values in rows 2 and 3 respectively. Inspecting Panel A, we find that the top performing fund generates an annualized alpha of 10.95%, being statistically significant in the parametric p-value evaluation. However, there is a stark difference with regards to the bootstrapped p-value being 0.55, leading us to not reject the null hypothesis. Inspecting the second and third top performers, we see a significant bootstrapped p-value, indicating that the fund managers indeed have enough skill to warrant their fees and add value to investors. This is a stark difference from the findings in Table 5, where the bootstrapped p-values of these percentiles were too high to warrant rejection of the null hypothesis.

Taking a look at the bottom performers ranked on alpha, we find significant negative alpha. In addition, when inspecting the corresponding bootstrapped p-value of the bottom performers, we clearly reject the null hypothesis when ranking on alpha. This holds except for the bottom performer, where the bootstrapped p-value suggests an insignificant value. These findings correspond to the findings of Table 5. Moving on to panel B in the table shows us that the top performers by no means generate abnormal returns warranting rejection of the null hypothesis. That is, the null of fund managers being able to cover their costs holds. Again, findings here correspond to the findings of applying the Kosowski et al. (2006) approach.

Inspecting the bottom performers tells a different story. The null hypothesis of fund managers being skillful enough to cover their costs is clearly rejected at all percentiles considered. Even at the bottom performing fund, contradicting the evidence from the Kosowski et al. (2006) approach in Table 5 where the bottom performer had a bootstrapped p-value of 0.19.

This implicates that the Norwegian mutual fund managers do not show statistically significant evidence of generating returns beyond covering their cost. Inspecting the top performers ranked on the t-statistic of alpha tells us that the generated returns of these funds clearly do not warrant rejection of the null. Contradicting the parametric p-value of the t-statistic of alpha.

Table 6: This table provides cross-sectional bootstrap results of all the Norwegian mutual funds in the sample 1987-2019, similar to Table 5, however, results are here based on applying the Fama and French (2010) procedure. Both panels show findings for the same top/bottom performers and percentiles. Panel A provides and ranks four-factor alpha estimates, along with bootstrapped p-values and parametric p-values of the corresponding percentiles. Panel B reports and ranks performance on the t-statistic of alpha, along with the bootstrapped p-values and parametric p-values. Columns, 1-5 report statistics of top funds, while column 6-10 reports for the bottom funds. The first row in the two panels reports estimated alpha and the t-statistic of alpha, respectively. Rows 2 and 3 report bootstrapped p-values and the parametric p-values of the fund in both panels, respectively. Statistics are based on 10,000 bootstrap resamples and ranked on their t-statistic of alpha in both panels.

	Top	2nd	3rd	Top 5%	Top 10%	Bottom 10%	Bottom 5%	3rd	2nd	Bottom
Panel A: Fund Ranked on Four-Factor Model Alpha										
Alpha	10.95	7.62	6.18	5.38	3.02	-7.01	-11.89	-16.93	-18.62	-19.06
Bootstrapped p-value	0.55	0.01	0.03	0.02	0.97	0.02	0.03	0.02	0.02	0.78
Parametric p-value	0.01	0.03	0.08	0.02	0.01	0.08	0.05	0.00	0.04	0.03
Panel B: Fund Ranked on Four-Factor Model t-statistics										
t-statistic	2.23	2.19	2.15	1.94	1.19	-1.73	-2.45	-2.96	-3.18	-3.22
Bootstrapped p-value	0.99	0.99	0.98	0.98	0.99	0.01	0.01	0.01	0.01	0.02
Parametric p-value	0.01	0.01	0.02	0.03	0.12	0.04	0.01	0.00	0.00	0.00

Through the application of the Fama and French (2010) method, we gather that outperforming funds are not skillful enough to generate returns above that of their fees. This rather, is due to luck. The funds which have the lowest generated returns, however, generate returns so low that they cannot be attributed to bad luck alone.

5.5 Harvey & Liu (2020) Double Bootstrap Approach

At the aggregate level, the previous section showed that the Norwegian mutual funds struggled to generate significant positive alpha. Now, we will apply the test procedure proposed by Harvey and Liu (2020), in order to draw further inference regarding the alpha generation of funds.

Figure 6 reports the distribution of t-statistics for the funds in our sample in panel A. Panel B present, the Receiver Operating Characteristic (ROC) curve of the fund data for positive t-statistic threshold. Finally, panel C presents the ROC curve of negative t-statistic thresholds for the underperformers. Specifically, for a given number of p_0 outperformers, the first-round bootstrap classifies the funds into true outperformers and zero-alpha performers. Then, the second round bootstrap is used to calculate the realized True Positive Rate (TPR) and False positive Rate (FPR) for each t-statistic cutoff, for each bootstrapped simulation.

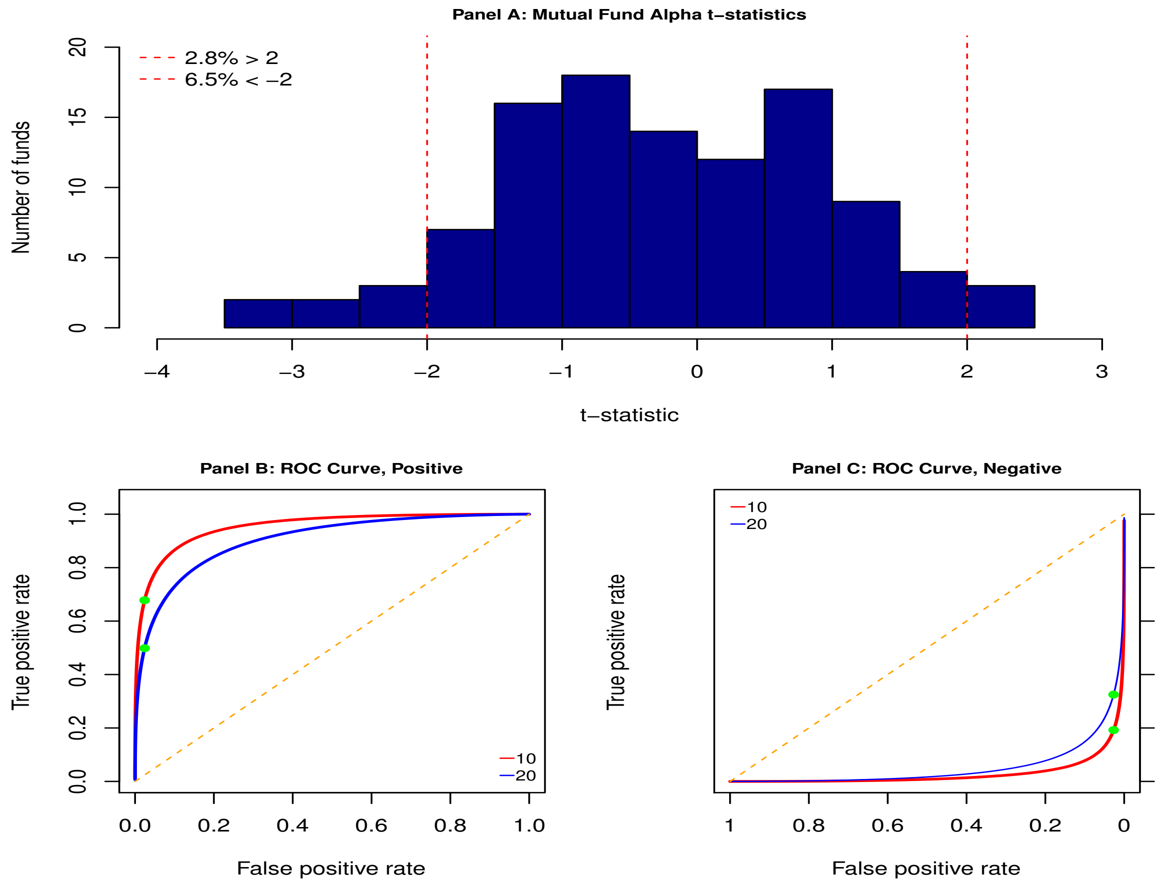
The ROC curve is an intuitive diagnostics plot, which is used to assess the performance of a classification method, e.g. a multiple-testing method (Harvey & Liu, 2020). Through the ROC curve, one can highlight the trade-off between FPR and TPR for different assumed levels of p_0. The closer the ROC curve is to the North-West corner of the plot the better, when investigating positive t-statistics, i.e., outperformance. That is, the closer a ROC curve is to the North-West corner, the better will the corresponding cutoff t-statistic be at separating true outperformance from false outperformance. The ROC curve is plotted for the two cases where $p_0 = 10$ and $p_0 = 20$.

As for underperformance, the further the ROC is to the South-East corner, the better. As the ideal would be a true positive rate of 1 and a false positive rate of 0. A 45-degree line has also been drawn through the plot, representing random classification.

We run $I = 100$ first-round bootstraps and $J = 10.000$ second round bootstraps, totaling 1.000.000 bootstraps for each fund. Lastly, we calculate the average TPR and FPR across simulations. The TPR is defined as the number of true rejections over the total number of true outperformers. The FPR is defined as the fraction of false rejections of the null to the total number of zero-alpha performers.

From Figure 6 Panel A, we see from the histogram that the distribution of actual t-statistics is skewed to the left, and we observe more t-statistics bellow -2 (6.5%) than above 2 (2.8%). Panel B shows us that the ROC curve of $p_0 = 10$ outperformers is closer to the North-West corner than the ROC curve for $p_0 = 20$. This is due to the fact that a smaller p_0 corresponds to a more select group, e.g., a higher average t-statistic for the outperformers and vice versa for underperformers. Finally, Panel C presents the ROC curve for $p_0 = 10$ and $p_0 = 20$ underperformers. We note that the ROC curves are drawn further to the South-East corner than the respective ROC curves are drawn to the North-West corner for outperformers in panel B. This indicates a higher signal-to-noise ratio for underperformers.

Figure 6: The figure presents both a t-statistic distribution (Panel A) and the Receiver Operating Characteristic (ROC) curve for both outperformers (Panel B) and underperformers (Panel C). Note that in the underperforming ROC curve plot, Panel C, the axis have been flipped. Using the Charhart four-factor model, t-statistics of alpha are estimated. The red lines (dashed) in Panel A represent t-statistics of -2 and 2. Points on the ROC curves correspond to a t-statistic of 2 in panel B, for outperformers, and -2 in Panel C, for underperformers.



We next investigate how the error rates change as we vary the cutoff t-statistic for a given number of p_0 out- and underperformers. We here compare the second round bootstrapped t-statistics, where we know which funds are generating abnormal performance, and which are generating zero-alpha performance. Through the use of equation (13) and equation (14), error rates are calculated. The Oratio is calculated as in equation (15). This is done for each cutoff t-statistic between 1.5 and 6 with 0.01 increments for outperformance, and between -1.5 and -6 with -0.01 increments for underperformance.

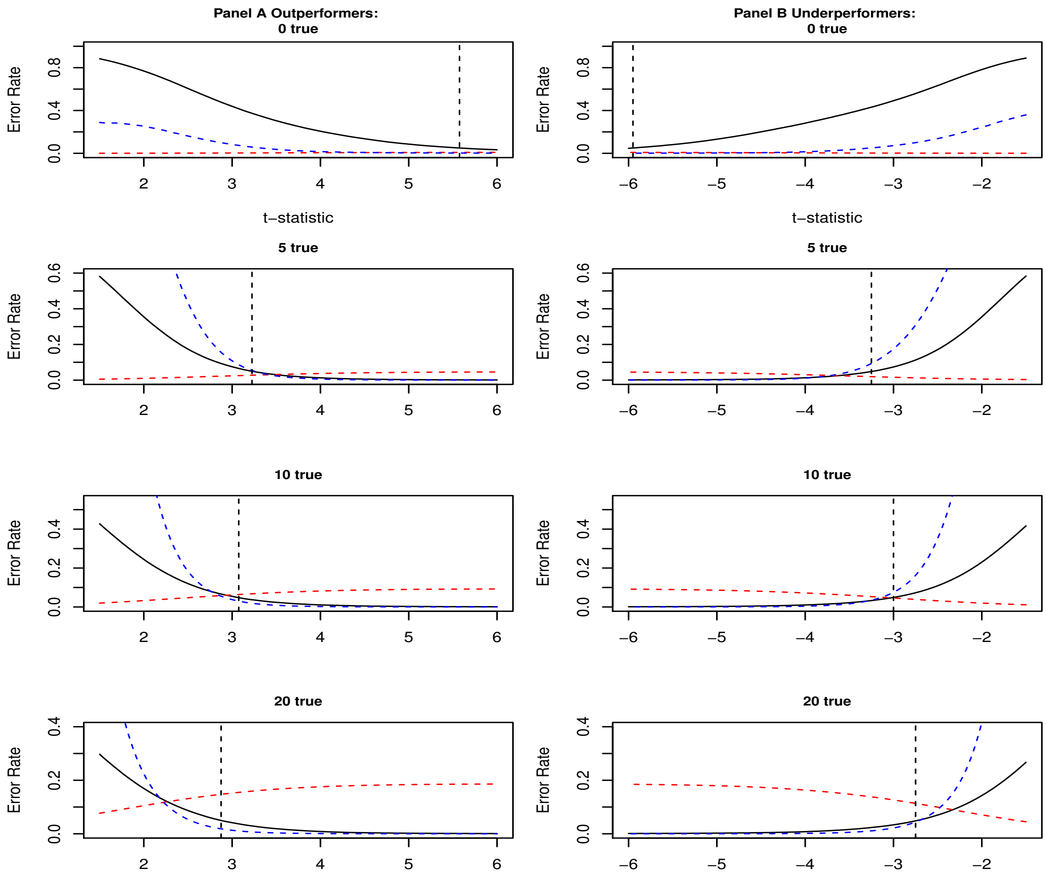
In Figure 7, Panel A shows Type I and Type II error rates, in addition to the odds ratio for the range of cutoff t-statistics for different levels of p_0: namely 0, 5, 10, and 20 funds outperforming. We see that the Type I error rate decreases as the statistical threshold increases in the panels A1-A4, for outperformers. However, as the Type I error rate decreases, the Type II error rate increases. This is the classical trade-off between Type I and Type II error rates extended to the multiple test. However, the increase in the Type II error rate is marginal when the cutoff t-statistic is increased. That is, the probability of failing to reject the null when the alternative is true remains relatively low.

The cutoff t-statistic corresponding to a Type I error rate of 5% is shown as the vertical lines in the plots. We note that the threshold t-statistic decreases as the number of outperformers increase. This follows from the fact that a higher number of true outperformers calls for a more lenient t-statistic threshold, in order to correctly reject the null hypothesis of no abnormal performance for the fund when more funds assumed to outperform.

In panels B1-B4, displaying error rates as a function of cutoff t-statistics for underperformers, we see the same pattern as for the outperformers in panels A1-A4. Note however that, as the statistical threshold decreases, the Type I error rate decreases. Further, the Type II error rate increases as the cutoff t-statistics are decreasing. The classic relationship between Type I and Type II error rates is shown once again. When the Type I error rate decreases, the Type II error rate increases.

In absolute values, we see that the cutoff t-statistics corresponding to a Type I error rate of 5% is very close between the outperformers and underperformers. However, we note that the Type II error rate is higher for the cutoff t-statistics measuring underperformance. Meaning that there is a greater possibility for a true underperforming fund manager not being detected than for a true outperformer not being detected.

Figure 7: The figure presents simulated error rates for both positive and negative t-statistics. Panel A show the simulated Type I (black) and Type II (red) error rates, and in addition odds ratio (blue) for a series of threshold t-statistics, ranging from 1.5 to 6.0 with 0.01 increments. This is conducted for $p_0 = 0, 5, 10,$ and 20. Panel B show the simulated Type I and Type II error rates of t-statistic thresholds of -1.5 to -6 with increments of -0.01. For each threshold t-statistic, the Type I and Type II error rate is calculated for each $p_0 = 0, 5, 10,$ and 20, number of assumed underperformers. In each panel, the cutoff t-statistic corresponding to a Type I error rate of 5% is represented by the vertical dashed line. Error rates are calculated based on $I = 100$ first-round bootstraps and $J = 10,000$ second-round bootstraps.



In Figure 8, we again see the pattern where the cutoff t-statistic corresponding to a Type I error rate of 5% decreases as the number of outperformers increase. From the plot, we can take away the fact that if one believes that zero outperforming funds are true, then the threshold t-statistic for the fund to overcome in order to reject the null hypothesis is 5.9 (if one wishes to control the Type I error rate to be 5%). Moving along the line, at $p_0 = 5$, we find a cutoff t-statistic of 3.23 giving the Type I error of 5%. Finally, for the highest assumed number of outperformers, we see that a t-statistic threshold of 2.88 would correspond to a 5% Type I error rate.

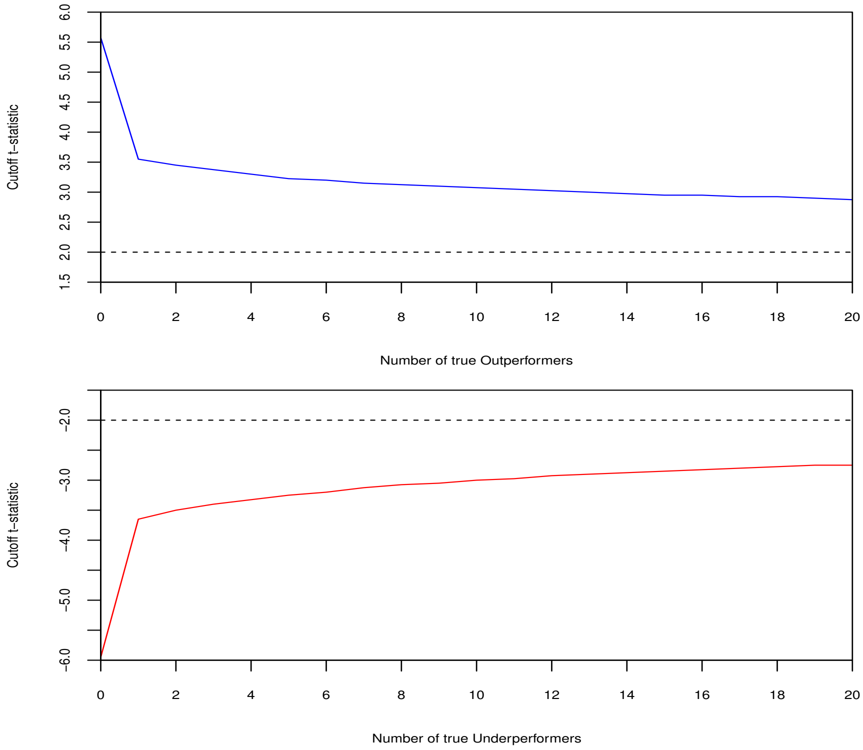
Inspecting the underperformers present a similar picture as for the outperformers. The threshold statistic starts out close to -6, and goes towards -2 as the number of underperformers increases. In the case where we have no true underperformers assumed, a threshold t-statistic of -5.95 suffices to hold the Type I error rate to the 5% level. The curve goes steeply upward to a cutoff t-statistic of -3.65 in the case of one true underperformer. Moving on along the spectrum of underperformers, the line gets less steep the more true underperformers there are assumed to be. For an assumed level of $p_0 = 10$ underperformers, we see a cutoff t-statistic of -3.0. Increasing the number of true underperformers up to 20 results in a cutoff t-statistic of -2.75, which is smaller in absolute terms than for the outperformers.

We can employ the findings in Figures 7 and 8 in the following manner: If one believes that the true number of outperforming funds is indeed 20, then, according to the findings in Figure 8 we should employ the statistical threshold of 2.88. In doing this, we know that there is only a 5% chance of picking a fund where the null is rejected that is actually not an outperformer. From Figure 7 we can draw the same conclusion, and we can also incorporate the cost of a Type I and a Type II error.

If one believes that the true number of outperformers is 10, and one may be willing to accept a Type I error that is higher than the usual level of 5%, say a 20% Type I error rate. That is, in the rejection of the null for 5 funds, one would expect one of these to be a zero-abnormal return performer. This is exactly the point of the proposed odds ratio, expressing the possibility of the outperformer being a true outperformer or not. Still, one is able to identify outperformers that may yield a return high enough to cover the cost of a zero alpha performer.

Table 7 shows cutoff t-statistics for different numbers of assumed out- and underperformers yielding an empirical Type I error rate of maximum 0.05 from the double bootstrap approach.

Figure 8: The figure presents cutoff t-statistics at a Type I error rate of 5%, as a function of p_0 out- and underperformers. The figure on top shows positive cutoff t-statistics for the p_0 assumed outperformers. The bottom figure shows the negative cutoff t-statistics corresponding to the p_0 assumed underperformers. The classical threshold t-statistic of 2 and -2 in hypothesis testing is shown by the horizontal (dashed) line in each plot respectively.



The corresponding Type II error rate is also presented, along with the number of rejections from the funds under consideration. From Panel A, we see that the cutoff t-statistic for $p_0 = 5$ is 3.23. The corresponding Type II error to this cutoff is 0.03, indicating that if we apply this cutoff, believing that 5 funds are truly outperforming, that the chance of falsely not declaring a true outperformer as outperforming is low. This gives a test power of 0.97 (= 1-0.03) in the Harvey and Liu (2020) framework.

We note that for the assumed levels of $p_0 = 10$ and $p_0 = 20$, the corresponding Type II error is above the 5% level that one usually wants in a hypothesis test. For instance, if we assume $p_0 = 20$ is true, then applying the 2.88 t-statistic threshold, we run the risk of not rejecting the null hypothesis when it is actually true in 15% of the cases. We do, however, see that using the statistical thresholds providing a Type I error rate of 5% for the right tail of the distribution yields zero rejections of the null for the funds in the sample.

From Panel B, showing the cutoff t-statistics for the left tail of the distribution, we see levels of the Type II error rate below 5% for the $p_0 = 0 to p_0 10$ underperformers, for the $p_0 = 20$

assumed underperformers, we find that the Type II error rate is at 11% when using a threshold t-statistic of -2.75. We do, however, for the underperformers, reject the null hypothesis in two occasions: when the assumed number of underperformers is 10, and when it is 20.

Finally, and most importantly, we note the last column in Table 7. The power of the tests conducted using the cutoff t-statistics. From calculating the Type II error rate in the usual way, as opposed to the Type II error rate calculation of Harvey and Liu (2020), we see that the test power is in fact very low. From this, we can conclude that the test would frequently fail to reject the null where the alternative is actually true.

Table 7: This table presents cutoff t-statistics for different values of p_0 out- and underperformers where the Type I error rate is required to be 5%. The first column shows the assumed number of outperformers and underperformers. Column 2 contain the cutoff t-statistic for the assumed number of out- and underperformers. The corresponding Type I, Type II error rates, and odds ratio are given in columns 3, 4, and 5 respectively. Column 6 shows the number of rejections of the null hypothesis for in the sample used. Finally, column 7 displays test power calculated using the usual and most frequent method of calculating Type II error rate in multiple tests. Panel A presents cutoff t-statistics for outperformers, whereas Panel B contains cutoff t-statistics for underperformers.

p_0	Cutoff	Type I	Type II	Odds ratio	n. rejections	Test Power
Panel A: Outperformers						
0	5.58	0.05	0.01	0	0	0.09
1	3.55	0.05	0.00	0.03	0	0.51
5	3.23	0.05	0.03	0.05	0	0.42
10	3.08	0.05	0.06	0.03	0	0.34
20	2.88	0.05	0.15	0.02	0	0.25
Panel B: Underperformers						
0	-5.95	0.05	0.01	0	0	0.10
1	-3.65	0.05	0.00	0.03	0	0.63
5	-3.25	0.05	0.02	0.09	0	0.59
10	-3.00	0.05	0.05	0.07	2	0.53
20	-2.75	0.05	0.11	0.04	3	0.44

From the analysis conducted using the proposed Harvey and Liu (2020) methodology, we find no evidence of top performers being skillful enough to generate abnormal returns net of fees for their customers. For the bottom performers, we find some evidence when the assumed number of underperformers is high. However, when calculating the Type II error rate for multiple test in the usual way, we see that the test power is too low. As such, we cannot rely on the results too heavily.

6 Conclusion

This thesis investigates whether Norwegian mutual fund managers are able to generate abnormal returns for their investors and whether this is due to skill or luck. We study 107 funds operating in the period 1987-2019, gathered from the TITLON database reporting net-of-fees returns for a sample free of survivorship bias. The four-factor model proposed by Carhart (1997) is applied as the primary performance model, both at aggregate and individual fund levels.

The bootstrapping approaches of Kosowski et al. (2006), Fama and French (2010), and Harvey and Liu (2020) have been applied, in order to distinguish between luck and skill. We take into consideration the Type I and Type II error rates in a multiple test. In addition, the bootstrap procedure of Kosowski et al. (2006) has been further applied in the evaluation of statistical significance through sensitivity analysis.

From aggregate fund performance results, we can conclude that Norwegian mutual fund managers on aggregate produce a net-of-fees non-significant annualized alpha of 0.04%. More specifically, this tells us that investors in the funds on aggregate do not receive a return after fees warranting the use of active management as opposed to passive index fund investments. This is in line with findings from other Scandinavian countries, as evident from the findings of Christensen (2013), Flam and Vestman (2017), and Korkeamaki and Smythe Jr. (2004). Neutral Aggregate fund performance is not a phenomena found only in Scandinavia. Researchers as early as Sharpe (1966) and Jensen (1968) concluded that active management was simply not worth the additional fees.

Individual fund parametric OLS regressions, tell us that funds indeed generate both positive- and negative significant abnormal returns. This however, holds true only amongst the top 7 funds, out of the 45 funds generating positive alpha ranked on the t-statistic of alpha in the sample. Whereas, the bottom 13 mutual funds ranked on the t-statistic of alpha out of the 62 funds generating negative alpha in the sample, have significant negative alpha at the 5% level. In addition, the bottom performers produce more extreme negative t-statistics than what we experience with the top performing funds. Indicating that the bottom performers drag the aggregate performance of the mutual funds down more than the outperformers are able to contribute with. Sharpe (1966) also found evidence of this among US mutual funds in the early stages of mutual fund performance evaluation. From the Swedish mutual fund market, Dahlquist et al. (2000) found that mutual funds indeed outperformed the market.

Further, stock-picking skill has been investigated through the use of the Kosowski et al. (2006) approach. Findings present minimal significant evidence of stock-picking skill amongst the 107 mutual funds investigated. Only in the application of the Fama and French (2010) approach, we find minimal evidence of outperformance when ranking funds on alpha. However, a lack of stock-picking skill is found amongst the fund managers generating the lowest abnormal returns net of fees. These findings contradict the findings of Gallefoss et al. (2015) in the Norwegian mutual fund market, in that they found stock picking skill amongst the mutual fund managers. Sørensen (2009) found no evidence of skill amongst the top performing Norwegian mutual funds in his study. However, the findings of this thesis and the studies of Gallefoss et al. (2015) and Sørensen (2009) all point to a lack of skill amongst the bottom performing mutual fund managers in Norway.

From the next applied approach of Fama and French (2010), in investigations of skill for the mutual funds. The adjusted bootstrap method applied, tells us that the null hypothesis of abnormal returns being due to luck is rejected. That is, conducting the same procedure as Sørensen (2009) in his investigation of Norwegian mutual funds, we find similar evidence for the sample used in this thesis, being of a longer time period than that investigated by Sørensen (2009). We do, however, find a glimmer of evidence in stock-picking skill among the top performing funds when calculating bootstrapped p-values and ranking funds based on alpha. Whereas bottom performing funds are not merely unlucky, they have a lack in stock-picking skills. Similar findings was presented by Flam and Vestman (2017) in his investigation in the Swedish mutual fund market. Additionally, Fama and French (2010) found no evidence of stock-picking skill in the US. mutual fund market among the top performing funds. Rather, they found, as seem evident from most studies conducting these bootstrapping procedures in investigation of mutual funds that the bottom performing funds had a lack of stock picking skill in order to cover their costs.

Through the double bootstrap approach of Harvey and Liu (2020) we find no evidence of skill amongst the top performing Norwegian mutual fund managers. Meaning that their abnormal return is due to luck rather than skill. When assuming a high amount of underperformers amongst the Norwegian mutual fund managers, cutoff t-statistics corresponding to a Type I error rate of 5% leads us to reject the null hypothesis for some bottom performers. Indicating that if the true number of underperformers is 20, then 3 of the managers in the sample have a lack of stock-picking skill in order to cover their cost. These findings are in line with the findings of Fama and French (2010) in their study of the US mutual fund market, where, as mentioned, the top performing mutual funds where not found to have stock-picking skill. We also contradict, as Fama and French (2010), the findings of Kosowski et al. (2006) in the US mutual fund market. Where the top performing funds where found to have stock picking skill.

In summary, all three approaches applied in the investigation of skill vs luck among the Norwegian mutual fund managers yield the same result. Stock-picking skill is not found among the top performing funds. There is as mentioned a glimmer of hope from the Fama and French (2010) where the second and third top, along with the fifth percentile performers generate bootstrapped t-statistics statistically significant. However, we cannot conclude that the fund managers exhibit stock-picking skill based on these results alone. Rather, with respect to the findings in this thesis as a whole. Performances is attributed to luck. For the bottom performers, we find, however, evidence of a lack of skill in all three approaches.

We must however express the power of the tests conducted here. Through the critique of Harvey and Liu (2020) we know that the test power is low for the bootstrapping procedures of Kosowski et al. (2006) and Fama and French (2010). In addition, when calculating the Type II error rate in the most usually applied way for multiple tests, we see that the power of the cutoff t-statistic methodology of Harvey and Liu (2020) is also too low. This implicates that the tests we apply in the investigation of skill vs. luck in mutual fund performance are not powerful enough. As such, we cannot exclude the possibility of stock-picking skill amongst the top performers.

References

Andrikogannopoulou, A., & Papakonstantinou, F. (2019). Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? *The Journal of Finance*, *74*(5), 2667-2688.

Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance*, *65*(1), 179-216.

Barras, L., Scaillet, O., & Wermers, R. (2019). *Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? a reply* (Working paper). McGill University.

Bauman, W. S., Conover, C. M., & Miller, R. E. (1998). Growth versus value and large-cap versus small-cap stocks in international markets. *Financial Analysts Journal*, *54*(2), 75-89.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300.

Bickel, P. J., & Freedman, D. A. (1984). Asymptotic Normality and the Bootstrap in Stratified Sampling. *The Annals of Statistics*, *12*(2), 470-482.

Blake, D., & Timmermann, A. (1998, 04). Mutual fund performance: Evidence from the UK. *European Finance Review*, *2*, 57-77.

Breeden, D. T., Gibbons, M. R., & Litzenberger, R. H. (1989). Empirical tests of the consumption-oriented CAPM. *The Journal of Finance*, *44*(2), 231-262.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, *5*(4), 553-580.

Busse, J. A., Goyal, A., & Wahal, S. (2010). Performance and persistence in institutional investment management. *The Journal of Finance*, *65*(2), 765-790.

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

Carpenter, J., & Lynch, A. W. (1999). Survivorship bias and attrition effects in measures of performance persistence. *Journal of Financial Economics*, *54*(3), 337-374.

Christensen, M. (2003). *Evaluating danish mutual fund performance* (WorkingPaper). Aarhus: The Aarhus School of Business.

Christensen, M. (2013). Danish mutual fund performance. *Applied Economics Letters*, *20*(8), 818-820.

Cuthbertson, K., Nitzsche, D., & O'Sullivan, N. (2008). UK mutual fund performance: Skill or luck? *Journal of Empirical Finance*, *15*(4), 613-634.

Dahlquist, M., Engström, S., & Söderlind, P. (2000). Performance and characteristics of swedish mutual funds. *The Journal of Financial and Quantitative Analysis*, *35*(3), 409-423.

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, *52*(3), 1035-1058.

Edelen, R. M. (1999). Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics*, *53*(3), 439-466.

Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). The persistence of risk-adjusted mutual fund performance. *The Journal of Business*, *69*(2), 133-157.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F., & French, K. R. (2010, 10). Luck versus skill in the cross section of mutual fund returns. *The Journal of Finance*, *65*, 1915 - 1947.

Ferson, W., & Yong, C. (2017). *How many good and bad fund managers are there really?* (Working Paper). California: University of Southern California.

Flam, H., & Vestman, R. (2014). Swedish equity mutual funds: Performance, persistence and presence of skill. *Swedish House of Finance*. (Reasearch Paper No. 14-04)

Flam, H., & Vestman, R. (2017). *Swedish equity mutual funds 1993-2013: Performance, persistence and presence of skill* (Tech. Rep.). CESifo. (Working Paper Series No. 6713)

Gallefoss, K., Hansen, H. H., Haukaas, E. S., & Molnár, P. (2015). What daily data can tell us about mutual funds: Evidence from norway. *Journal of Banking & Finance*, *55*, 117-129. (Global Governance and Financial Stability)

Genovese, C., & Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64*(3), 499–517.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, *70*(3), 393-408.

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, *51*(3), 783-810.

Hall, P., Horowitz, J. L., & Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, *82*(3), 561-574.

Hall, P., & Martin, M. A. (1988). On bootstrap resampling and iteration. *Biometrika*, *75*(4), 661-671.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics. *The Journal of Finance*, *72*(4), 1399-1440.

Harvey, C. R., & Liu, Y. (2020). False (and missed) discoveries in financial economics. *The Journal of Finance*, *75*(5), 2503-2553.

Harvey, C. R., Liu, Y., & Zhu, H. (2016, 10). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5-68.

Horowitz, J. L. (2003). Bootstrap methods for markov processes. *Econometrica*, *71*(4), 1049-1082.

Ioannidis, J. P. A. (2005, 08). Why most published research findings are false. *PLOS Medicine*, *2*(8), null.

Ippolito, R. A. (1989, 02). Efficiency with costly information: A study of mutual fund performance, 1965–1984. *The Quarterly Journal of Economics*, *104*(1), 1-23.

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

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, *23*(2), 389-416.

Korkeamaki, T. P., & Smythe Jr., T. I. (2004). Effects of market segmentation and bank concentration on mutual fund expenses and returns: Evidence from Finland. *European Financial Management*, *10*(3), 413-438.

Kosowski, R., Timmermann, A., White, H., & Wermers, R. (2006, 02). Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis. *The Journal of Finance*, *61*, 2551-2595.

Lehmann, B. N., & Modest, D. M. (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *The Journal of Finance*, *42*(2), 233-265.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13-37.

Lo, A. W., & MacKinlay, A. C. (1990). Data-snooping biases in tests of financial asset pricing models. *The Review of Financial Studies*, *3*(3), 431-467.

Malkiel, B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance*, *50*(2), 549-572.

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

Moskowitz, T. J. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion. *The Journal of Finance*, *55*(4), 1695-1703.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*(4), 768-783.

Ødegaard, B. A. (2020). *Asset pricing data at OSE* (Database). Stavanger: University of Stavanger.

Ødegaard, B. A. (2021a, 03). Empirics of the Oslo Stock Exchange. Asset pricing results. 1980-2020. *University of Stavanger*.

Ødegaard, B. A. (2021b, 03). Empirics of the Oslo Stock Exchange. Basic, descriptive, results. *University of Stavanger*.

Otten, R., & Bams, D. (2002). European mutual fund performance. *European Financial Management*, *8*(1), 75-101.

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

Reinganum, M. R. (1981). A new empirical perspective on the CAPM. *The Journal of Financial and Quantitative Analysis*, *16*(4), 439-462.

Sarkar, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *The Annals of Statistics*, *34*(1), 394–415.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, *19*(3), 425-442.

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

Sørensen, L. (2009, 10). Mutual fund performance at the oslo stock exchange. *SSRN Electronic Journal*.

Titlon, U. o. T. (2020). *Titlon - financial data for norwegian academic institutions* (Tech. Rep.). Tromsø: University of Tromsø.

Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance*, *55*(4), 1655-1695.

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

Yan, X., & Zheng, L. (2017, 04). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *Review of Financial Studies*, *30*, 1382-1423.

Appendix A

Table A.I: The table presents the number of observations we have for each fund in column 2. Further, annualized mean return is displayed in column 3 and standard deviation is shown in column 4. The third and fourth moments, kurtosis and skew, for each funds net returns is presented in column 5 and 6, respectively. Lastly, the maximum and minimum monthly net return is presented in the penultimate and rightmost columns, respectively

Fund	Obs	Mean	Std.dev	Kurtosis	Skewness	Max	Min
ABIF Norge ++	56	8.02	23.41	2.53	-0.31	13.53	-16.25
Alfred Berg Aksjef Norge	115	10.36	21.14	4.88	-0.77	13.06	-24.97
Alfred Berg Aksjespar	106	9.30	22.90	5.17	-0.87	13.34	-27.99
Alfred Berg Aktiv	289	14.11	22.62	5.59	-0.77	21.08	-27.05
Alfred Berg Aktiv II	182	9.27	25.30	4.23	-0.60	17.89	-27.37
Alfred Berg Gambak	350	15.25	22.68	5.69	-0.40	28.50	-27.38
Alfred Berg Humanfond	241	10.40	20.27	6.04	-0.97	16.12	-25.88
Alfred Berg N. Pensjon	52	13.14	21.02	7.56	-1.36	11.91	-24.80
Alfred Berg Norge	147	10.62	24.59	4.90	-0.96	17.10	-27.01
Alfred Berg Norge +_gml	197	11.07	23.57	5.23	-0.97	17.13	-26.91
Alfred Berg Norge Classic	351	10.70	20.99	6.04	-1.06	17.10	-27.01
Alfred Berg Norge Etisk	146	11.83	23.99	5.58	-1.04	16.65	-27.84
Alfred Berg Norge Inst	72	12.12	9.64	4.26	-0.95	6.81	-7.98
Alfred Berg Vekst	72	8.89	26.53	4.89	-0.51	19.33	-27.82
Arctic Norwegian Equities Class A	109	9.52	11.36	4.51	-0.67	9.49	-9.26
Arctic Norwegian Equities Class B	110	10.02	12.01	4.44	-0.57	9.79	-9.20
Arctic Norwegian Equities Class D	83	11.79	9.41	4.55	-0.92	7.11	-8.51
Arctic Norwegian Equities Class I	110	10.10	11.92	4.41	-0.58	9.65	-9.19
Atlas Norge	263	10.53	24.20	6.67	-0.09	36.85	-25.25
Banco Norge	38	13.25	24.05	2.73	-0.33	13.89	-17.12
C WorldWide Norge	294	13.40	20.51	6.07	-0.89	19.80	-27.52
Carnegie Aksje Norge	210	14.47	23.37	5.06	-0.86	19.80	-27.52
Danske Invest Aktiv Formuesf. A	21	16.69	15.54	3.59	-0.83	7.59	-10.67
Danske Invest Norge Aksj. Inst 1	237	11.94	19.51	5.69	-0.93	15.46	-22.85
Danske Invest Norge Aksj. Inst 2	158	10.30	18.72	7.29	-1.16	15.04	-22.73
Danske Invest Norge I	312	11.17	20.09	6.63	-1.03	14.85	-28.80
Danske Invest Norge II	312	11.95	20.18	6.64	-1.02	14.91	-29.49
Danske Invest Norge Vekst	312	15.34	22.27	9.61	0.33	41.77	-25.68
Delphi Norge	307	15.13	23.39	5.07	-0.54	23.01	-24.93
Delphi Vekst	193	11.02	26.07	4.03	-0.33	25.54	-23.04
DNB Norge	289	9.30	20.10	5.38	-0.84	15.81	-24.12
DNB Norge (Avanse I)	327	10.90	21.89	5.10	-0.96	15.96	-26.42
DNB Norge (Avanse II)	287	9.65	21.57	5.39	-0.96	16.05	-26.40
DNB Norge (I)	295	11.48	24.59	18.28	1.31	59.30	-24.16
DNB Norge (III)	283	10.72	20.25	5.37	-0.87	15.87	-24.17
DNB Norge (IV)	206	13.95	19.48	5.95	-0.89	15.97	-24.24
DNB Norge R	12	15.86	9.99	3.70	-1.20	4.48	-5.54
DNB Norge Selektiv (II)	214	11.99	20.04	5.37	-0.77	16.85	-23.74
DNB Norge Selektiv (III)	307	11.44	20.05	5.22	-0.82	16.99	-24.07
DnB Real-Vekst	157	6.14	30.41	27.68	2.15	68.90	-40.26
DNB SMB	226	14.13	23.76	4.17	-0.47	17.48	-26.49
Eika Norge	196	14.43	19.27	6.96	-1.02	18.40	-24.93
Eika SMB	187	8.34	23.44	4.28	-0.67	17.06	-22.94
FIRST Generator	112	13.14	19.55	4.41	-0.77	15.51	-18.90
FIRST Norge Fokus	14	11.50	10.02	3.70	-0.99	5.22	-6.09
Fokus Barnespar	32	-0.65	27.06	6.44	-1.21	12.75	-28.09
Fondsfinans Aktiv II	48	-0.60	23.15	2.93	-0.23	14.34	-16.48
Fondsfinans Norge	205	16.90	20.17	5.66	-0.78	16.32	-25.73
FORTE Norge	107	10.00	14.50	4.05	-0.08	14.49	-11.60

Fund	Obs	Mean	Std.dev	Kurtosis	Skewness	Max	Min
FORTE Trønder	81	16.37	11.97	3.27	-0.14	9.46	-8.80
GAMBAK Oppkjøp	19	3.50	19.03	3.48	0.35	13.94	-9.16
GJENSIDIGE AksjeSpar	152	10.54	22.73	5.18	-0.94	16.59	-26.70
GJENSIDIGE Invest	104	15.24	20.35	5.39	-0.85	13.34	-21.18
Globus Aktiv	88	15.14	29.39	3.35	-0.30	23.56	-22.63
Globus Norge	103	7.71	29.32	3.38	-0.35	22.34	-23.36
Globus Norge II	95	11.98	28.55	3.40	-0.24	23.12	-22.91
Handelsbanken Norge	300	12.34	20.67	7.08	-1.17	17.75	-28.82
Handelsbanken Norge A10	18	3.12	12.61	3.52	-1.18	4.93	-8.50
Holberg Norge	229	11.86	19.79	4.70	-0.52	15.94	-23.90
K-IPA Aksjefond	37	11.49	22.96	5.08	-0.97	12.32	-21.75
KLP Aksjeinvest	97	5.38	21.14	4.67	-0.78	14.92	-22.21
KLP AksjeNorge	250	11.52	20.36	6.24	-0.91	17.59	-29.77
Landkreditt Norge	122	7.14	20.38	5.14	-0.74	17.13	-20.70
Landkreditt Utbytte	83	13.84	7.31	3.44	-0.76	4.65	-4.68
Landkreditt Utbytte I	19	10.44	7.35	2.71	-0.39	4.21	-3.82
NB-Aksjefond	207	9.97	22.50	5.14	-0.94	18.24	-24.78
Nordea Avkastning	396	11.81	21.95	5.70	-0.87	20.68	-27.57
Nordea Barnespar	47	-2.20	21.16	2.68	-0.36	11.38	-16.35
Nordea Kapital	298	12.70	20.23	5.92	-1.01	16.70	-25.72
Nordea Kapital II	84	14.40	22.67	2.80	-0.47	13.37	-17.50
Nordea Kapital III	70	12.74	23.25	2.75	-0.56	13.33	-17.46
Nordea Norge Pluss	105	10.10	13.32	4.17	-0.65	12.10	-11.09
Nordea Norge Verdi	287	12.47	18.99	5.66	-0.87	15.17	-24.46
Nordea SMB	213	6.80	23.64	3.54	-0.23	18.26	-23.23
Nordea SMB II	70	-13.70	26.47	3.13	0.17	18.70	-19.14
Nordea Vekst	337	10.62	23.00	4.85	-0.85	19.50	-26.22
ODIN Norge	331	15.35	20.88	5.42	-0.43	22.78	-24.09
ODIN Norge A	50	10.85	9.76	4.55	-1.19	4.69	-8.58
ODIN Norge B	50	10.60	9.78	4.56	-1.20	4.66	-8.63
ODIN Norge D	50	10.61	9.77	4.56	-1.20	4.67	-8.61
ODIN Norge II	139	12.24	19.43	6.11	-0.99	13.59	-23.98
Orkla Finans 30	162	18.56	21.83	4.48	-0.71	14.70	-26.16
Pareto Aksje Norge	220	13.62	18.55	6.53	-0.84	16.11	-26.09
PLUSS Aksje (Fondsforval)	277	11.29	20.58	5.32	-0.72	17.56	-25.51
PLUSS Markedsverdi (Fondsforv)	300	11.75	19.46	6.21	-0.98	15.95	-25.03
Postbanken Aksjevekst	97	7.37	23.68	3.15	-0.40	14.84	-19.72
RF Aksjefond	116	10.31	21.45	4.27	-0.73	13.50	-23.83
RF-Plussfond	54	16.89	24.89	2.47	-0.36	14.45	-17.05
Sbanken Framgang Sammen	47	12.73	9.46	3.81	-0.73	6.72	-7.18
SEB Norge LU	67	-4.99	25.29	4.24	-0.65	15.62	-26.07
Skandia Horisont	97	11.63	22.30	4.13	-0.76	16.23	-21.52
Skandia SMB Norge	97	0.71	23.81	5.44	-1.01	13.83	-27.32
SR-Bank Norge A	12	16.50	10.17	2.50	-0.49	5.24	-4.64
SR-Bank Norge B	12	16.50	10.16	2.50	-0.49	5.24	-4.64
Storebrand Aksje Innland	282	11.33	20.13	6.08	-1.02	15.39	-26.50
Storebrand AksjeSpar	226	6.75	15.06	4.30	-0.89	10.32	-14.04
Storebrand Norge	396	12.76	21.54	5.48	-0.89	17.30	-28.83
Storebrand Norge A	43	23.16	24.49	2.76	-0.52	14.64	-17.17
Storebrand Norge Fossilfri	33	11.18	7.01	4.35	-0.89	4.51	-5.34
Storebrand Norge I	237	11.35	20.56	6.21	-1.00	14.85	-28.59
Storebrand Norge Institusjon	39	8.48	14.62	3.65	-0.53	9.89	-9.72
Storebrand Optima Norge	221	11.84	21.07	6.04	-0.99	14.59	-29.29
Storebrand Vekst	328	15.77	24.02	6.70	0.01	36.71	-30.06
Storebrand Verdi	265	11.05	20.03	6.18	-0.97	13.50	-26.53
Storebrand Verdi N	22	7.06	10.48	2.85	-0.53	5.92	-5.55
Terra Norge	187	9.47	24.69	4.46	-0.75	18.81	-26.20
VÅR Aksjefond	39	8.57	24.43	6.49	-1.19	11.46	-26.08

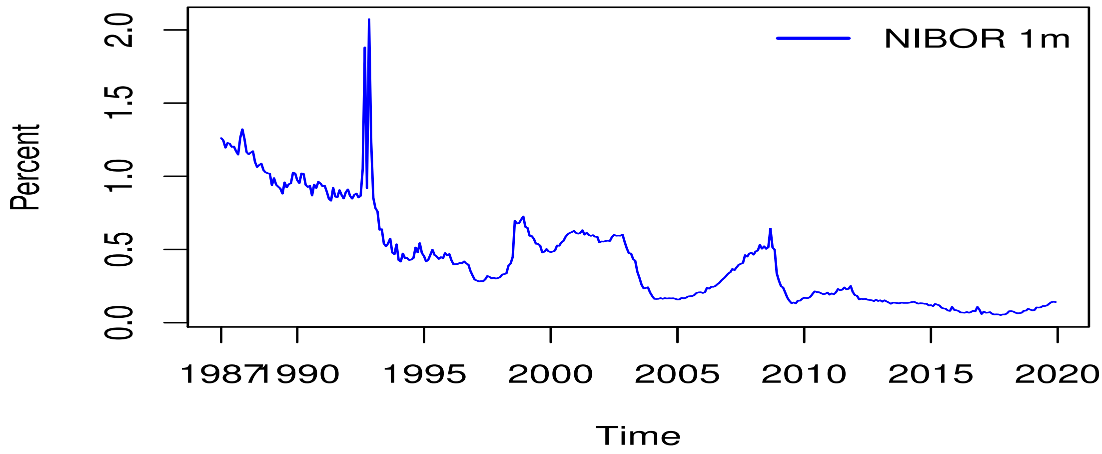
Table A.II: This table presents factor loadings of the Charhart four-factor model and adjusted R^2 for each individual fund. OLS regression has been run in order to estimate the factor loadings, where the full period of a funds return series has been utilized. Columns 2 to 5 present factor loadings, whereas column 6 present the adjusted R^2 for the net excess return of the fund.

Fund	β_{MKT}	β_{SMB}	β_{HML}	β_{PR1YR}	$adj - R^2$
ABIF Norge ++	1.04	-0.07	-0.05	-0.11	0.96
Alfred Berg Aksjef Norge	1.01	0.07	-0.00	-0.04	0.95
Alfred Berg Aksjespar	1.07	0.11	-0.01	0.02	0.92
Alfred Berg Aktiv	1.14	0.27	-0.18	0.03	0.86
Alfred Berg Aktiv II	1.09	0.31	-0.18	-0.06	0.87
Alfred Berg Gambak	1.10	0.31	-0.27	0.11	0.80
Alfred Berg Humanfond	0.99	-0.01	-0.09	-0.07	0.91
Alfred Berg N. Pensjon	1.07	0.07	-0.03	-0.01	0.93
Alfred Berg Norge	1.06	0.03	-0.08	-0.09	0.95
Alfred Berg Norge +_gml	1.05	0.05	-0.08	-0.06	0.95
Alfred Berg Norge Classic	1.05	0.02	-0.04	-0.01	0.94
Alfred Berg Norge Etisk	1.05	0.03	-0.16	-0.14	0.94
Alfred Berg Norge Inst	0.84	-0.02	-0.08	0.12	0.84
Alfred Berg Vekst	1.16	0.31	-0.11	0.16	0.80
Arctic Norwegian Equities Class A	0.84	0.05	-0.10	0.17	0.75
Arctic Norwegian Equities Class B	0.92	0.06	-0.10	0.18	0.80
Arctic Norwegian Equities Class D	0.82	0.03	-0.08	0.18	0.76
Arctic Norwegian Equities Class I	0.92	0.06	-0.10	0.18	0.80
Atlas Norge	1.13	0.14	-0.27	-0.02	0.85
Banco Norge	1.05	0.13	-0.18	-0.19	0.92
C WorldWide Norge	1.01	-0.02	-0.16	0.03	0.92
Carnegie Aksje Norge	1.02	0.00	-0.16	0.00	0.93
Danske Invest Aktiv Formuesf. A	0.78	0.37	0.39	0.33	0.69
Danske Invest Norge Aksj. Inst 1	0.96	-0.04	-0.03	-0.09	0.92
Danske Invest Norge Aksj. Inst 2	0.96	-0.04	-0.02	-0.03	0.91
Danske Invest Norge I	0.99	0.00	-0.05	-0.10	0.90
Danske Invest Norge II	1.00	0.01	-0.04	-0.10	0.91
Danske Invest Norge Vekst	1.09	0.41	-0.23	0.02	0.77
Delphi Norge	1.16	0.30	-0.23	-0.03	0.84
Delphi Vekst	1.11	0.37	-0.29	-0.09	0.84
DNB Norge	1.00	-0.05	-0.03	-0.06	0.96
DNB Norge (Avanse I)	0.94	0.00	-0.05	-0.09	0.92
DNB Norge (Avanse II)	0.98	-0.01	-0.06	-0.07	0.94
DNB Norge (I)	1.00	0.05	-0.02	-0.04	0.74
DNB Norge (III)	1.01	-0.03	-0.04	-0.06	0.96
DNB Norge (IV)	1.01	-0.02	-0.07	-0.05	0.96
DNB Norge R	0.86	0.33	0.12	-0.14	0.93
DNB Norge Selektiv (II)	1.01	-0.03	-0.06	-0.06	0.95
DNB Norge Selektiv (III)	1.03	0.05	-0.07	-0.05	0.94
DnB Real-Vekst	0.99	0.05	-0.04	-0.01	0.46
DNB SMB	1.19	0.47	-0.12	-0.19	0.79
Eika Norge	1.03	0.13	-0.04	-0.10	0.89
Eika SMB	0.99	0.18	-0.04	-0.22	0.85
FIRST Generator	1.46	0.33	-0.07	0.00	0.73
FIRST Norge Fokus	0.73	-0.33	-0.19	-0.03	0.77
Fokus Barnespar	0.97	0.04	-0.02	-0.27	0.83
Fondsfinans Aktiv II	0.96	-0.07	-0.00	-0.17	0.90
Fondsfinans Norge	1.03	0.07	-0.06	-0.12	0.87
FORTE Norge	1.06	0.07	-0.09	0.03	0.70
FORTE Trønder	0.80	0.03	-0.04	0.05	0.47
GAMBAK Oppkjøp	0.49	0.26	-0.14	0.41	0.68
GJENSIDIGE AksjeSpar	0.95	0.05	0.03	0.00	0.93
GJENSIDIGE Invest	1.00	0.16	0.09	0.02	0.94

Fund	β_{MKT}	β_{SMB}	β_{HML}	β_{PR1YR}	$adj-R^2$
Globus Aktiv	1.18	0.23	-0.21	-0.32	0.82
Globus Norge	1.16	0.28	-0.21	-0.35	0.84
Globus Norge II	1.17	0.25	-0.22	-0.33	0.81
Handelsbanken Norge	1.03	0.01	-0.07	-0.01	0.90
Handelsbanken Norge A10	1.04	-0.09	-0.08	0.18	0.90
Holberg Norge	1.00	0.24	-0.10	-0.10	0.84
K-IPA Aksjefond	0.97	0.16	0.10	0.00	0.85
KLP Aksjeinvest	0.95	0.01	-0.04	-0.10	0.90
KLP AksjeNorge	1.02	-0.02	-0.05	-0.06	0.92
Landkreditt Norge	0.95	0.11	-0.01	-0.15	0.83
Landkreditt Utbytte	0.53	0.07	0.03	0.08	0.50
Landkreditt Utbytte I	0.53	0.16	0.03	0.01	0.60
NB-Aksjefond	1.00	0.07	-0.02	-0.16	0.92
Nordea Avkastning	0.96	-0.01	-0.07	-0.07	0.83
Nordea Barnespar	0.98	-0.04	-0.05	-0.01	0.92
Nordea Kapital	1.02	0.02	-0.08	-0.06	0.93
Nordea Kapital II	1.03	-0.12	-0.07	-0.10	0.92
Nordea Kapital III	1.03	-0.01	-0.08	-0.17	0.94
Nordea Norge Pluss	1.07	0.12	-0.08	0.03	0.84
Nordea Norge Verdi	0.94	0.16	-0.04	-0.12	0.86
Nordea SMB	1.10	0.52	-0.09	-0.15	0.83
Nordea SMB II	1.04	0.55	-0.13	-0.09	0.78
Nordea Vekst	1.01	0.03	-0.04	-0.07	0.91
ODIN Norge	1.00	0.29	0.06	-0.09	0.79
ODIN Norge A	0.81	0.01	-0.05	-0.02	0.77
ODIN Norge B	0.81	0.01	-0.05	-0.02	0.77
ODIN Norge D	0.81	0.01	-0.05	-0.02	0.77
ODIN Norge II	0.98	0.30	-0.05	-0.06	0.82
Orkla Finans 30	1.05	0.14	-0.05	-0.09	0.91
Pareto Aksje Norge	0.95	0.18	0.00	-0.04	0.84
PLUSS Aksje (Fondsforval)	0.97	-0.02	-0.08	-0.08	0.89
PLUSS Markedsverdi (Fondsforv)	0.95	-0.11	-0.04	-0.07	0.94
Postbanken Aksjevekst	1.02	0.05	-0.18	-0.10	0.92
RF Aksjefond	0.96	0.02	-0.06	-0.13	0.92
RF-Plussfond	1.11	0.20	-0.30	-0.19	0.87
Sbanken Framgang Sammen	0.89	-0.04	-0.04	0.07	0.84
SEB Norge LU	1.07	0.03	-0.08	-0.06	0.92
Skandia Horisont	1.05	0.21	-0.09	0.03	0.86
Skandia SMB Norge	1.05	0.43	-0.13	-0.12	0.82
SR-Bank Norge A	1.30	0.27	0.03	0.16	0.90
SR-Bank Norge B	1.30	0.27	0.03	0.16	0.90
Storebrand Aksje Innland	1.01	-0.04	-0.04	-0.03	0.97
Storebrand AksjeSpar	0.65	0.06	-0.14	-0.10	0.72
Storebrand Norge	0.98	0.00	-0.04	-0.05	0.89
Storebrand Norge A	1.06	0.03	-0.13	-0.20	0.92
Storebrand Norge Fossilfri	0.47	-0.10	-0.11	0.03	0.55
Storebrand Norge I	1.04	0.01	-0.06	-0.09	0.94
Storebrand Norge Institusjon	0.98	0.03	-0.07	0.05	0.91
Storebrand Optima Norge	1.04	0.02	-0.06	-0.10	0.92
Storebrand Vekst	1.05	0.25	-0.40	-0.03	0.71
Storebrand Verdi	0.99	-0.05	0.11	0.01	0.94
Storebrand Verdi N	0.84	0.02	0.07	-0.07	0.93
Terra Norge	1.06	0.12	-0.16	-0.10	0.92
VÅR Aksjefond	1.11	0.09	0.21	0.06	0.90

Appendix B

Figure B.I: This figure presents the Norwegian interbank offered rate (NIBOR) one month rate, used as a proxy for the risk free rate for the full sample period, 1987-2019. The rate is given in percent.



Appendix C

Table C.I: The table presents annualized alpha from four-factor model in percent for each individual fund. Alpha estimates are displayed in column 1. In addition, parametric t-statistic and the corresponding parametric p-value is presented in the third and fourth column, respectvely. Lastly, the bootstrapped p-value of the respective fund is presented in column 5.

Fund	Alpha	t-statistic	Parametric p-value	bootstrapped p-value
Danske Invest Norge Aksj. Inst 1	3.02	2.23	0.01	0.80
FIRST Norge Fokus	10.95	2.19	0.01	0.52
Landkreditt Utbytte	5.38	2.15	0.02	0.30
Fondsfinans Norge	3.78	1.94	0.03	0.34
Danske Invest Norge Aksj. Inst 2	3.26	1.94	0.03	0.17
PLUSS Markedsverdi (Fondsforv)	1.95	1.83	0.03	0.16
Landkreditt Utbytte I	7.62	1.82	0.03	0.09
FORTE Trønder	6.18	1.43	0.08	0.56
Storebrand Norge I	1.60	1.36	0.09	0.58
Storebrand Norge Fossilfri	4.18	1.24	0.11	0.70
Storebrand Optima Norge	1.72	1.19	0.12	0.70
Landkreditt Norge	3.19	1.15	0.12	0.66
Carnegie Aksje Norge	1.78	1.15	0.13	0.56
DNB Norge R	6.11	1.11	0.13	0.52
Danske Invest Norge II	1.39	1.07	0.14	0.53
Storebrand Verdi N	2.32	1.02	0.15	0.55
Storebrand Verdi	1.13	0.99	0.16	0.52
Eika Norge	1.71	0.97	0.17	0.47
Pareto Aksje Norge	1.80	0.95	0.17	0.41
K-IPA Aksjefond	4.72	0.90	0.18	0.46
Nordea Norge Verdi	1.35	0.88	0.19	0.41
Alfred Berg Norge	1.37	0.85	0.20	0.41
PLUSS Aksje (Fondsforval)	1.25	0.82	0.21	0.39
Nordea Kapital	0.93	0.81	0.21	0.34
Storebrand Norge	0.97	0.74	0.23	0.43
C WorldWide Norge	0.92	0.74	0.23	0.37
DNB SMB	1.91	0.71	0.24	0.36
Alfred Berg Norge Inst	1.30	0.66	0.26	0.43
Holberg Norge	1.15	0.59	0.28	0.55
ODIN Norge	1.07	0.55	0.29	0.59
Danske Invest Norge I	0.72	0.55	0.29	0.50
ODIN Norge A	1.55	0.54	0.29	0.44
DNB Norge Selektiv (II)	0.58	0.50	0.31	0.48
ODIN Norge D	1.31	0.45	0.32	0.54
ODIN Norge B	1.28	0.45	0.33	0.49
VÅR Aksjefond	1.91	0.42	0.34	0.48
Storebrand Aksje Innland	0.23	0.30	0.38	0.79
Nordea Avkastning	0.48	0.29	0.38	0.74
ABIF Norge ++	0.70	0.29	0.39	0.68
Alfred Berg Humanfond	0.31	0.21	0.42	0.82
DNB Norge (IV)	0.17	0.16	0.44	0.87
KLP AksjeNorge	0.21	0.15	0.44	0.85
Skandia Horisont	0.26	0.08	0.47	0.92
Handelsbanken Norge	0.08	0.06	0.48	0.92
Storebrand Vekst	0.06	0.02	0.49	0.93
Alfred Berg Norge +_gml	-0.01	-0.01	0.50	0.06
DNB Norge (I)	-0.19	-0.07	0.47	0.03
DNB Norge (III)	-0.06	-0.07	0.47	0.04
Storebrand AksjeSpar	-0.20	-0.10	0.46	0.04
Delphi Norge	-0.23	-0.12	0.45	0.04

Fund	Alpha	t-statistic	Parametric p-value	Bootstrapped p-value
ODIN Norge II	-0.39	-0.15	0.44	0.04
Alfred Berg Gambak	-0.38	-0.19	0.42	0.02
Nordea Kapital II	-0.71	-0.26	0.40	0.01
FORTE Norge	-1.04	-0.33	0.37	0.00
Danske Invest Norge Vekst	-0.74	-0.33	0.37	0.00
DnB Real-Vekst	-2.30	-0.36	0.36	0.00
DNB Norge Selektiv (III)	-0.38	-0.36	0.36	0.01
Nordea Norge Pluss	-0.81	-0.38	0.35	0.01
Sbanken Framgang Sammen	-1.01	-0.41	0.34	0.01
Banco Norge	-2.11	-0.50	0.31	0.00
Eika SMB	-1.23	-0.52	0.30	0.00
SEB Norge LU	-1.68	-0.53	0.30	0.00
Storebrand Norge A	-2.27	-0.56	0.29	0.00
Fondsfinans Aktiv II	-2.49	-0.65	0.26	0.00
Alfred Berg Norge Etisk	-1.23	-0.68	0.25	0.00
Arctic Norwegian Equities Class D	-1.54	-0.69	0.25	0.00
Terra Norge	-1.28	-0.69	0.24	0.00
Delphi Vekst	-1.99	-0.74	0.23	0.00
DNB Norge (Avanse I)	-1.00	-0.80	0.21	0.00
Handelsbanken Norge A10	-3.21	-0.84	0.20	0.00
NB-Aksjefond	-1.36	-0.85	0.20	0.00
Atlas Norge	-1.91	-0.91	0.18	0.00
FIRST Generator	-3.58	-0.91	0.18	0.00
Alfred Berg Aktiv	-1.71	-0.92	0.18	0.00
Alfred Berg Norge Classic	-0.94	-0.93	0.18	0.00
Nordea Barnespar	-3.02	-0.94	0.17	0.00
KLP Aksjeinvest	-2.44	-0.97	0.17	0.00
SR-Bank Norge B	-7.07	-1.05	0.15	0.00
SR-Bank Norge A	-7.08	-1.05	0.15	0.00
Orkla Finans 30	-2.07	-1.06	0.14	0.00
RF Aksjefond	-2.32	-1.09	0.14	0.00
Postbanken Aksjevekst	-2.97	-1.17	0.12	0.00
Alfred Berg Aktiv II	-2.93	-1.20	0.12	0.00
Globus Aktiv	-6.13	-1.21	0.11	0.00
Nordea Kapital III	-3.23	-1.24	0.11	0.00
Alfred Berg N. Pensjon	-3.49	-1.24	0.11	0.00
Arctic Norwegian Equities Class A	-2.92	-1.30	0.10	0.00
Nordea Vekst	-1.77	-1.31	0.10	0.00
Storebrand Norge Institusjon	-3.56	-1.33	0.09	0.00
Arctic Norwegian Equities Class I	-2.86	-1.37	0.09	0.00
Arctic Norwegian Equities Class B	-2.99	-1.42	0.08	0.00
RF-Plussfond	-7.01	-1.43	0.08	0.01
DNB Norge	-1.21	-1.44	0.07	0.02
DNB Norge (Avanse II)	-1.72	-1.53	0.06	0.01
Fokus Barnespar	-11.89	-1.67	0.05	0.00
Alfred Berg Vekst	-8.95	-1.72	0.04	0.00
Globus Norge II	-8.34	-1.73	0.04	0.01
GAMBAK Oppkjøp	-18.62	-1.81	0.04	0.01
Danske Invest Aktiv Formuesf. A	-19.06	-1.89	0.03	0.01
Globus Norge	-8.54	-1.99	0.02	0.01
Alfred Berg Aksjef Norge	-3.32	-2.03	0.02	0.02
Alfred Berg Aksjespar	-4.98	-2.14	0.02	0.02
GJENSIDIGE AksjeSpar	-4.28	-2.45	0.01	0.01
Nordea SMB	-6.44	-2.63	0.00	0.01
GJENSIDIGE Invest	-5.31	-2.96	0.00	0.01
Nordea SMB II	-16.93	-3.18	0.00	0.02
Skandia SMB Norge	-12.25	-3.22	0.00	0.19

Appendix D

Table D.I: This table presents bootstrapped p-values where the block length used in the bootstrapping of residuals has been altered from the initial block length of 4 used for the main part of this thesis. Funds are ranked on t-statistic of alpha, Bootstrapped- and parametric p-values of the funds are presented in the following rows. Panel A displays results using a block length of 1, whereas panel B and C presents results of applying a block length of 2 and 10 respectively. Columns 2-6 present findings for top percentiles, bottom percentiles results are presented in columns 7-11.

	Top	2.nd	3.rd	Top 5%	Top 10%	Bottom 10%	Bottom 5%	3.rd	2.nd	Bottom
Panel A: Block Length of 1										
t-statistic	2.23	2.19	2.15	1.94	1.24	-1.81	-2.45	-2.96	-3.18	-3.22
Bootstrapped p	0.82	0.55	0.32	0.21	0.76	0.01	0.00	0.00	0.02	0.17
Parametric p	0.01	0.01	0.02	0.03	0.11	0.04	0.01	0.00	0.00	0.00
Panel B: Block Length of 2										
t-statistic	2.23	2.19	2.15	1.94	1.24	-1.81	-2.45	-2.96	-3.18	-3.22
Bootstrapped p	0.81	0.53	0.32	0.2	0.73	0.01	0.00	0.01	0.02	0.18
Parametric p	0.01	0.01	0.02	0.03	0.11	0.04	0.01	0.00	0.00	0.00
Panel c: Block Length of 10										
t-statistic	2.23	2.19	2.15	1.94	1.24	-1.81	-2.45	-2.96	-3.18	-3.22
Bootstrapped p	0.78	0.48	0.25	0.14	0.61	0.01	0.00	0.00	0.02	0.18
Parametric p	0.01	0.01	0.02	0.03	0.11	0.04	0.01	0.00	0.00	0.00

Table D.II: This table presents bootstrapped p-values where the minimum number of observations in order to be included in the sample has been altered. Panel A displays the baseline bootstrap used in this thesis. Panels B-D results for when a minimum of 26, 36, and 60 is used in fund selection. Funds are ranked on t-statistic of alpha, Bootstrapped- and parametric p-values of the funds are presented in the following rows 2 and 3 for each panel. Columns 2-6 present findings for top percentiles, bottom percentiles results are presented in columns 7-11.

	Top	2.nd	3.rd	Top 5th	Top 10th	Bottom 10th	Bottom 5th	3.rd	2.nd	Bottom
Panel A: Baseline bootstrap, minimum 12 obs										
t-statistic	2.23	2.19	2.15	1.94	1.19	-1.73	-2.45	-2.96	-3.18	-3.22
bootstrapped p-value	0.80	0.52	0.29	0.17	0.70	0.01	0.00	0.00	0.02	0.19
p-value	0.01	0.01	0.02	0.03	0.12	0.04	0.01	0.00	0.00	0.00
Panel B: Minimum 24 observations										
t-statistic	2.23	2.15	1.94	1.83	1.15	-1.72	-2.45	-2.96	-3.18	-3.22
bootstrapped p	0.75	0.50	0.48	0.24	0.76	0.01	0.00	0.00	0.01	0.16
parametric p	0.01	0.02	0.03	0.03	0.12	0.04	0.01	0.00	0.00	0.00
Panel C: Minimum 36 observations										
t-statistic	2.23	2.15	1.94	1.83	1.15	-1.72	-2.45	-2.96	-3.18	-3.22
Bootstrapped p	0.74	0.48	0.47	0.23	0.74	0.01	0.00	0.00	0.01	0.16
parametric p	0.01	0.02	0.03	0.03	0.13	0.04	0.01	0.00	0.00	0.00
Panel D: Minimum 60 observations										
t-statistic	2.23	2.15	1.94	1.83	1.15	-1.72	-2.45	-2.96	-3.18	-3.22
bootstrapped p	0.70	0.43	0.40	0.16	0.59	0.01	0.00	0.00	0.01	0.15
parametric p	0.01	0.02	0.03	0.03	0.13	0.04	0.01	0.00	0.00	0.00

Table D.III: This table presents bootstrapped p-values of portfolios of funds. The average t-statistic is calculated for the top 1 to n funds, and compared to the average bootstrapped t-statistics for that same number of top funds. The same procedure is used for the bottom funds. The average t-statistic of the top 1 to n funds is given in row 1, and the corresponding bootstrapped p-value is presented in row 2.

	Top	Top 2	Top 3	Top 5	Top 10	Bottom 10	Bottom 5	Bottom 3	Bottom 2	Bottom
Average t	2.23	2.21	2.19	2.09	1.81	-2.43	-2.89	-3.12	-3.20	-3.22
Bootstrapped p	0.80	0.71	0.61	0.50	0.44	0.01	0.01	0.03	0.07	0.19

Table D.IV: This table presents bootstrapped p-values of the Kosowski et al. (2006) method using a block length of 4 for the sub-periods investigated in this thesis. Funds are ranked on parametric t-statistic of alpha in all panels. Panel A present the bootstrap results of the period as a whole. Panel B to panel E present bootstrap results for each sub period respectively. Each panel present parametric t-statistic, bootstrapped p-value and parametric p-value in row 1-3 respectively. Top fund results is presented in column 2-6 in descending order respectively. Results for bottom performers is shown in the 7 to 11 column, where column 6 present the bottom 10th percentile and column 11 present results of the bottom performer.

	Top	2.nd	3.rd	5%	10%	10%	5%	3.rd	2.nd	Bottom
Panel A:Fund Ranked on Four-Factor Alpha t-statistic, full period (1987 - 2019)										
t-alpha	2.23	2.19	2.15	1.94	1.24	-1.81	-2.45	-2.96	-3.18	-3.22
Bootstrapped p	0.80	0.52	0.30	0.17	0.70	0.01	0.01	0.01	0.02	0.19
parametric p	0.01	0.01	0.02	0.03	0.11	0.04	0.01	0.00	0.00	0.00
Panel B:Funds Ranked on Four-Factor Alpha t-statistic (1987 Q1 - 1995 Q1)										
t-alpha	1.57	1.02	0.46	0.06	-0.70	-1.08	-1.31	-3.12	-3.16	-3.28
Bootstrapped p	0.66	0.77	0.97	0.98	0.02	0.00	0.01	0.00	0.02	0.23
Parametric p	0.06	0.15	0.32	0.48	0.24	0.14	0.09	0.00	0.00	0.00
Panel C:Funds Ranked on Four-Factor Alpha t-statistic (1995 Q2 - 2003 Q2)										
t-alpha	2.46	1.20	1.19	1.08	0.28	-2.12	-2.64	-3.16	-3.18	-4.03
Bootstrapped p	0.34	0.99	0.98	0.95	1.00	0.00	0.00	0.00	0.01	0.02
Parametric p	0.01	0.12	0.12	0.14	0.39	0.02	0.00	0.00	0.00	0.00
Panel D:Funds Ranked on Four-Factor Alpha t-statistic (2003 Q3 - 2011 Q3)										
t-alpha	1.91	1.77	1.64	1.40	1.13	-1.33	-1.89	-2.44	-2.84	-3.12
Bootstrapped p	0.90	0.80	0.73	0.70	0.42	0.00	0.00	0.00	0.00	0.00
Parametric p	0.03	0.04	0.05	0.08	0.13	0.09	0.03	0.01	0.00	0.00
Panel B:Funds Ranked on Four-Factor Alpha t-statistic (2011 Q4 - 2019 Q4)										
t-alpha	2.19	2.15	1.82	1.55	1.02	0.93	0.66	0.61	0.61	0.60
Bootstrapped p	0.73	0.41	0.53	0.51	0.74	0.68	0.94	0.84	0.76	0.69
Parametric p	0.01	0.02	0.03	0.06	0.15	0.18	0.26	0.27	0.27	0.27

Discussion paper, master thesis International

Kristian Fjærbu Løhaugen Vikstøl

The topic of the thesis ''Are Active Norwegian Mutual Fund Managers Paid for Luck or Skill", utilizes the method of Kosowski, Timmermann, White, and Wermers (2006) and Fama and French (2010) in the investigation of skill or luck amongst the Norwegian mutual fund managers. Further, Harvey and Liu (2020) proposed a new approach in this sphere, also applying the bootstrapping methodology. Bootstrapping methodology is key in the research conducted, in both skill and multiple test correction used in the thesis. The subject of whether mutual fund managers are skillful is heavily studied abroad.

The thesis examines a total of 107 Norwegian mutual funds, in the period extending from 1987 to 2019. In the sample, both the funds still open up to 2019, and the funds that have closed before this are included in the sample. The null hypothesis states that there is no true performance, meaning that mutual fund managers are not able to generate significant outperformance nor underperformance in relation to the risk factors. Kosowski et al. (2006) and Fama and French (2010) method is replicated, as these two methods are highly recognized in distinguishing luck from skill amongst mutual fund managers, generating p-values or, equivalently, likelihoods as to whether the manager is in possession of skill. Through the application, we find no evidence of skill amongst the mutual fund managers at a 5% significance level.

However, there is evidence of having a lack of skill amongst the worst performers in Norwegian mutual funds. These findings are in line with the findings of Kosowski et al. (2006) and Fama and French (2010) as they also found significant evidence of a lack of skill. Fama and French, however, found little evidence of skill amongst the managers, where Kosowski et al. (2006) found skill. From the investigation applying the method of Harvey and Liu (2020) we did not find evidence of out- or underperformance. That said, applying their method is relying on the signal to noise in the data, where a low signal to noise could result in the findings being less relevant.

In relation to the concept “international” it is clear that the methods applied are readily adopted by researchers around the world. Both the methods proposed by Kosowski et al. (2006) and Fama and French (2010) has been applied by multiple researchers. In addition, the statistical method of bootstrapping, is also recognized as a well-functioning statistical method. The funds under investigation are comprised of Norwegian stocks, in addition, they are required to always hold at least 80\% Norwegian equities. These funds thus operate mainly in Norway, however, they are definitely exposed to international trends. Stocks listed on the Oslo Stock exchange are to a large degree internationally affected both in terms of international investors buying and selling the stock, in addition to relations to countries abroad. The international trend of sustainability is hitting the financial market of Norway too. ESG investing introduces constraints on the stocks available for the funds following the ESG standards. This is not necessarily a constraint which imposes a problem for mutual fund managers in generating positive net returns for their investors. That is, international and national institutions are contributing to ESG investing through different channels, making the investments more attractive for funds and investors.

Funds are required to take action on these trends if they are to generate value for their investors as companies aimed at creating a sustainable future have generated high returns in the past. Being part of an international financial world also warrants the funds to follow international requirements for financial services. With MiFID II introduced, funds are required to act accordingly in order to be recognized as a responsible financial actor.

In addition, the thesis investigates mutual fund performance both on individual and aggregate levels through factor models. These models have been applied to a large degree in international studies since the inception in 1968 where Jensen proposed a single factor model, measuring the abnormal performance of a funds through its alpha (Jensen, 1968). Further, Fama and French (1993) and Carhart (1997) extended the single factor model by additional risk factors taking into consideration market anomalies.

These factor models have been presented to me at multiple occasions throughout the financial master program. Both in statistical analysis and in subjects concerning financial theory and portfolio management. I find that these models contribute to the understanding of how the financial markets work, and point to important aspects of the financial markets.

Through the use of factor model's researchers have struggled to warrant the use of active mutual fund management as opposed to investing in passive index funds. Several Scandinavian studies conclude that the mutual fund managers are not able to do much more than collect fees paid by investors (Christensen, 2003; Flam and Vestman, 2017; Korkemaki and Smythe Jr, 2004). There seems, however, to be multiple findings of active managers which are able to generate positive net-of-fees returns for their investors. The main issue here, is that these are relatively few in comparison with active managers which actually underperform in relation to their benchmark.

In relation to international research, investigators, as mentioned, are struggling to find evidence as to warrant the use of active mutual funds. Despite this fact, we see, as depicted in the thesis, that the assets under management for Norwegian mutual funds have been growing throughout the years after 2000. This points to a disagreement between researchers and mutual fund investors. Researchers suggest that active investment is not worthwhile, whereas investors believe that active investment adds value. There are many possible explanations as to account for this difference in beliefs, and a discussion around this would be outside the scope of the discussion paper. It is worth noting, however, that there seems to be a disagreement between these two groups.

There has also been a trend towards the use of cheap index funds as an alternative to actively managed funds. Index funds are supposed to follow the index of some sector, stock exchange, or the likes. Here, researchers and investors are in agreement. Researchers have argued that any attempt to ''beat the market" is a foolish attempt (Fama, 1970). One aspect of this discussion which I have struggled to find, is the viewpoint where if you invest in a passively managed index fund, you are guaranteed to lose against the market when fees are deducted. Although these fees are low in relation to the fees collected by active mutual funds, they are still guaranteeing loss if the passive index is indeed able to replicate the returns of the given index it is supposed to follow.

Fama and French (2010) also argues that if any fund is able to beat the market, they do so with a negative impact on another fund, that is, in sum, the funds do not beat the market. This argument expresses the need for investors to not pick funds at will, but search for funds that may outperform in relation to others. We also see from the cumulative returns in the thesis that an equally weighted portfolio of funds that are still open as of December 2019 has a higher cumulative return than that of the benchmark and funds that have been closed throughout the sample. The returns have not been tested for significance, but we do however note this difference.

There has also been a debate amongst researchers internationally as to what benchmark one should use in the evaluation of mutual funds. In an ideal world, one would apply the market portfolio from

CAPM, however, such a portfolio is not readily available and one needs to use a proxy instead. From the discussions of multiple researchers, one should use a broad index, which is suited to represent a benchmark for the active funds at hand. That is, in order to evaluate fund performance in line with the consensus with international research standards on the topic, one cannot make use of one benchmark for all studies. Meaning that if we are investigating mutual funds in Norway, we have to incorporate a benchmark which represent the market where the Norwegian mutual funds invest. We also here have to take into consideration the stocks which are available to the funds. If we are investigating funds which have the opportunity to buy small and risky stocks, these should also be represented in the benchmark used for comparison. A point made by Elton, Gruber, and Blake (1996) in their critique of Ippolito (1989).

Through the use of Harvey and Liu (2020) approach, published in one of the forefront financial journals, we might be in the early stages of an up and coming statistical method for evaluating funds, financial assets, and financial strategies. Taking into consideration the statistical power of tests is important, yet, many papers fail to consider this. When Harvey and Liu evaluated the approach applied by Fama and French (2010), they found that the Type II error rate of the test was too high, resulting in low test power. From this, the tests used to evaluate skill in funds might have to be adjusted in order to generate a high enough statistical power to be certain in inferences.

In the past few years, statistical power has been up and coming in the discussion amongst researchers (Harvey and Liu, 2020). Statistical power is of great importance, as inferences may be misleading if the results of the study are affected by the Type I error rate and Type II error rate of the test conducted. The significance of error rates was presented already at the first statistics course. This affects the degree of which we can rely on the findings. Harvey and Liu (2020) argued in their paper that the focus have, in too large a degree, been on generating statistically significant evidence, in order to be published.

From this, I conclude that the active mutual fund managers, and the methods used are definitely internationally orientated. The Norwegian mutual fund managers invest in stocks that are not in a closed environment, as such, managers need to pay attention to international trends. In addition, they are also not the only ones who have the ability to invest in these funds. The methods has been applied by several researchers, in different parts of the world. As to the findings of the thesis, the investigation leads us to say that we find evidence of unskilled managers. From this, we can say that findings are in line with the findings from other studies in other countries investigating the same or similar hypothesis.

References

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

Christensen, M. (2013). Danish mutual fund performance. *Applied Economics Letters, 20*(8), 818-820.

Elton, E. J., Gruber, M. J., & Blake, C. R. (1996). The persistence of risk-adjusted mutual fund performance. *The Journal of Business, 69*(2), 133-157.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383-417.

Fama, E. F., & French, K. R. (2010, 10). Luck versus skill in the cross section of mutual fund returns. *The Journal of Finance, 65*, 1915 - 1947.

Flam, H., & Vestman, R. (2014). Swedish equity mutual funds: Performance, persistence and presence of skill. *Swedish House of Finance.* (Reasearch Paper No. 14-04)

Harvey, C. R., & Liu, Y. (2020). False (and missed) discoveries in financial economics. *The Journal of Finance, 75*(5), 2503-2553.

Ippolito, R. A. (1989, 02). Efficiency with costly information: A study of mutual fund performance, 1965-1984. *The Quarterly Journal of Economics, 104*(1), 1-23.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance, 23*(2), 389-416.

Korkeamaki, T. P., & Smythe Jr., T. I. (2004). Effects of market segmentation and bank concentration on mutual fund expenses and returns: Evidence from Finland. *European Financial Management, 10*(3), 413-438.

Kosowski, R., Timmermann, A., White, H., & Wermers, R. (2006, 02). Can mutual fund "stars" really pick stocks? new evidence from a bootstrap analysis. *The Journal of Finance*, *61*, 2551-2595.