Norwegian School of Economics
Bergen, Autumn, 2020

Persistence of Mutual Funds Returns

Do Norwegian portfolio managers consistently beat market benchmarks?

Akram Salekroshani and Elmira Vagapova

Supervisor: Jørgen Haug

Master Thesis, MSc, Finance

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

We conduct a study on 99 actively managed Norwegian mutual funds from 1996 to 2019 to investigate whether funds deliver returns in excess of passive benchmarks and if the funds' performance persists over time. We use the Fama and French 3-factor and Carhart 4-factor models as proxies for the passive benchmarks. Additionally, we bootstrap the results using Fama and French methodology to differentiate skill from luck in the mutual funds' returns. We find no evidence of skill among managers to produce superior returns for investors. To test for performance persistence, we employ recursive portfolio approach, construct contingency tables, and obtain cross-product ratios with corresponding Z and Chi-Squared statistics. The results of persistence tests suggest that only past losers remain losers in the subsequent period, while past winners are more likely to switch from outperforming to underperforming in the following period. Our study is the most comprehensive and up to date analysis of actively managed Norwegian mutual funds, which can be of interest to researchers and investors.

Acknowledgements

We would like to express our gratitude to our supervisor, professor Jørgen Haug, who provided us with invaluable insights and supported our efforts throughout the whole working process. The feedback that we received helped us to stay focused and enhance our work.

We are also thankful to the IT department at the Norwegian School of Economics for giving us access to Morningstar Direct so that we can download and use the data in our thesis.

Last but not least, we are grateful to Bernt Arne Ødergaard, professor and researcher at the University of Stavanger, who constructed risk pricing factors for Fama & French (1996) and Carhart (1997) benchmark models and gave free access to those factors on his website.

Contents

List of Tables

List of Figures

1. Introduction

We examine the performance and performance persistence of the survivorship bias-free sample of 99 actively managed Norwegian equity mutual funds from 1996 to 2019. We address two main questions: i) Do actively managed Norwegian mutual funds outperform the passive benchmark net of fees? ii) Can the actively managed funds perform persistently?

Our focus is on actively managed funds since these funds claim to persistently generate an excess return above the passive benchmark through superior stock selection or market timing skills. Seeing the substantial growth in assets under management of mutual funds in Norway, we find the study on actively managed mutual funds is valuable due to its implications for investors and their choice of investment.

Numerous studies have been conducted to provide insight into whether actively managed mutual funds gain abnormal returns. Most of the researchers like Jensen (1968), Elton (1993), Carhart (1997), and Fama and French (2010) discovered that mutual fund managers not only fail to beat the market but also underperform the passive benchmark on net returns. The researchers found that only a handful of mutual funds can outperform the passive benchmark on net returns through the manager's stock-picking ability. In the study that was carried out on the Norwegian mutual funds market, Sørensen (2010) concluded that there is no evidence of outperformance for the actively managed funds.

Considering the primary methodologies used in the literature, we employ the most efficient methods to investigate Norwegian mutual funds' performance. We employ the Fama and French 3-factor and Carhart 4-factor models as the benchmarks assuming that an alpha estimation obtained above the benchmark models is the excess return gained by managers' skills. However, due to non-normality in the distribution of mutual funds returns, it is challenging to distinguish whether excess returns originate from managerial ability or luck. To differentiate between luck and stock-picking ability, we bootstrap alphas by employing Fama and French bootstrap method (2010). The findings indicate that the best-performing funds fail to beat the market, while the worst-performing funds generate significantly negative alphas when funds are ranked by their alphas. When funds are ranked by t-statistics of alphas, the bootstrapped results imply no abnormal performance among Norwegian mutual funds.

Since some studies found evidence for positive and negative abnormal performance, the topic of how long that performance persists became a subject of interest for both investors and academic researchers. By analyzing performance persistence, it is easier to differentiate skill from luck since the longer the performance lasts, the higher the chances that performance is not a noisy measurement but an indication of superior or inferior managerial performance. If mutual funds outperform the market consistently, we can conclude that managerial skill, not luck, creates excess return. On the contrary, if mutual funds underperform the market continuously, it cannot be only due to bad luck, and it is an indication of weak managerial performance.

Similar to the performance research, the main studies on performance persistence have been conducted on the US market data. The findings in the studies indicate different degrees of performance persistence for various time horizons. While Malkeil (1995) found no evidence that mutual funds repeat their performance over time, Hendricks, Petal, and Zechhauser (1993) and Carhart (1997) argued that there is persistence among the worst-performing mutual funds.

We also investigate whether Norwegian actively managed mutual funds persistently produce an abnormal return against the passive indices. We conduct the highly relevant and valid persistence tests found in the literature: recursive portfolio test, contingency table, and cross-product ratio. The tests are conducted using both net raw returns and risk-adjusted returns to get a more comprehensive insight into the persistence level in inferior and superior funds' returns. We analyze performance persistence from a short to a long-time window to see if persistence varies with time horizon. The results of recursive portfolio tests are bootstrapped to identify if winners remain winners by luck or due to managerial skill. The bootstrapping technique also aims to detect if losers continue the lousy performance due to bad luck or the manager's lack of skill. This analysis reveals that past-losers in the lowest-ranked quantiles continue producing negative risk-adjusted returns both in the short and long term. In comparison, past winners show either zero alpha or underperform in the following period. The results of the bootstrap test indicate inferior managerial skills rather than bad luck. The evidence of contingency tables and cross-product ratio using raw and risk-adjusted returns shows that the results are consistent with the previous ones in which losers persist in being losers in the following periods. At the same time, winners do not experience persistent positive performance over time. Besides, the worst-performing funds have the highest likelihood to disappear in the subsequent period.

The findings of our study contribute to the literature as the most up to date analysis of actively managed Norwegian mutual funds, in which we employ a variety of proven techniques to evaluate performance and performance persistence. This research also sheds light on the Norwegian mutual funds' performance against passive indices, which provides useful material for investors in the equity markets to make a more informed investment decision. If actively managed funds cannot create excess returns above the passive benchmark net of fees, investors would be better off by following low-cost passive investment strategies. The findings are also significant from the economic perspective because financial intermediaries are well-rewarded due to their value-adding activities. If the superior performance is created by luck rather than managerial ability, paying the management fees and investing in mutual funds seems irrational and value destructive for the economy (Berk & Green, 2004).

The study proceeds as follows: section 2 provides a literature review of the most significant and relevant research on the topic, section 3 describes the data set and the variables that are employed in the research, section 4 defines the techniques and models which are implemented in the study, section 5 provides the empirical results of performance measurements as well as the finding of the performance persistence tests and section 6 presents the conclusion.

2. Literature Review

The ability of mutual fund managers to persistently generate returns has been a subject of numerous studies. The literature on the topic has been evolving over the years, and it has started with the question of how to measure mutual fund performance appropriately.

Jensen (1968) used the alpha variable to evaluate the performance of 115 open-end mutual funds in the United States for the period 1945 to 1964. He defined alpha as a return that is generated in excess of the CAPM model benchmark. Statistically significant nonzero alpha implies managerial ability to forecast security prices. If a manager is successful in forecasting, the alpha is positive. The alpha is negative if a manager performs poorly in forecasting. The results of the study reveal that, on average, funds are not able to predict future security prices that outperform buy-the-market-and-hold policy even with the returns gross of management expenses. Jensen was one of the first researchers to conclude that the mutual funds cannot systematically outperform the benchmark model.

Grinnblatt and Titman (1992) introduced a new performance benchmark constructed from passive portfolios to account for the biases that CAPM benchmark exhibited. For instance, small or income-oriented funds seem to outperform using the conventional CAPM benchmark. Their new benchmark had additional risk factors that controlled for the firm characteristics, such as size and past returns, to evaluate fund performance more accurately. Grinnblatt and Titman examined the performance of 279 US mutual funds from 1974 to 1984 and conducted a 5-year persistence test. The results indicate positive performance persistence, which cannot be explained by benchmark inefficiencies related to yield, CAPM beta, interest rate sensitivity, or skewness. The persistent performance could have been partially explained by persistent differences in transaction costs and fees.

Hendricks, Patel, and Zeckhauser (1993) also conducted a performance persistence test but on the shorter horizon. They used quarterly free-survivorship bias data (net of management fees) from 1974 to 1988, including open-end, no-load, growth-oriented equity funds. The results indicate that performance persistence is significant for the best and worst-performing funds up to four quarters. However, the outperformance of top funds is marginal in comparison to the market benchmarks.

In order to further examine persistent returns of the best and worst-performing funds, Jegadeesh and Titman (1993) tested the strategy of selling losers and buying winners from 1965 to 1989. They established that the winners' positive returns minus losers' portfolios continue for one year and vanishes entirely within two years after formation.

The short-lived performance persistence detected in the studies was commonly attributed to the managerial ability to beat the market. Malkiel (1995) tested that claim by conducting a study on all the funds that have ever existed. He used annual mutual fund returns from 1971 to 1991 and concluded that funds underperform the market not only net of management fees but also gross of all expenses excluding load fees. Malkiel argued that many studies tend to use data sets that contain only survived funds; therefore, the results incur strong survivorship bias, which distorts the conclusion on the performance of mutual funds. Although he documented performance persistence, he suggested that it can result from survivorship bias since he included mutual funds that existed for at least two years. Malkiel concluded that there is no compelling evidence for the superior managerial skill to generate strong, persistent excess returns.

A similar study was done by Brown and Goetzmann in 1995. They conducted a persistence test on an annual basis with survivorship-bias-free data and found evidence for performance persistence in excess of ex-ante benchmarks. They found out that persistence varies with the time horizon chosen in the studies. Brown and Goetzmann also noticed irregular reversed patterns in the winners and losers' performance that is not captured by the risk-adjusted models and cannot be related to managers' strategies. Additionally, the weak-performing funds are more likely to disappear but not entirely; thus, the survived funds might have driven the performance persistence results for the worst-performing funds.

Carhart (1997) found evidence for unexplained short-term persistence only among the worst-performing funds. He argued that the short-term persistence found by Hendricks et al. (1993) is explained by the Fama and French 3-factor model, momentum factor, differences in load fees, expense ratios, and portfolio turnover, but not by the superior stock-picking skills of the managers. Carhart concluded that there is little evidence of managerial skill, which nevertheless disappears with higher fees for investors.

Bollen and Busse (2004) modified the Carhart approach to test performance persistence. They ranked mutual funds by the risk-adjusted quarter returns over the following three-month period. They found that the top decile generates a statistically significant abnormal return. The results they got are robust with changes in market timing, stock selection, market models, and with inclusion of momentum strategy. The results clash with Carhart's results, which Bollen and Busse attributed to their particular use of risk-adjusted returns and change in the duration of ranking and post-ranking periods. Despite the robust results of superior performance, Bollen and Busse questioned the economic significance of their findings.

While performance persistence was evidenced in various studies, the economic significance of this finding became a subject of particular interest. Kosowski, Timmermann, Wermers, and White (2006) investigated whether the excess returns are attributed to luck or funds' manager skills in selecting stocks. They used the bootstrapping technique to differentiate between luck and stock-picking skills. The empirical findings on the US mutual funds between 1975 and 2002 indicate that mutual funds' performance cannot be solely explained by luck.

Kacperczyk, Sialm, and Zheng (2008) examined if the unobserved actions might have explained superior persistent fund performance. They included factors like timing of trades, related transaction costs, managers' informational advantage, trading costs, the agency problem, and investor externalities in their model. Kacperczyk, Sialm, and Zheng studied the effect of these unobserved actions by analyzing the "return gap" between the investor and the buy-and-hold returns. The results show that the effect of unobserved factors is persistent in the long run both for the bottom and the top-performing funds; however, the impact of these actions varies significantly across funds. They also argued that the return gap can be a significant "predictive power" for fund performance.

To find whether alphas are genuine or the result of luck, Barras, Scaillet, and Wermers (2010) employed the False Discovery Rate (FDR) measure. By accounting for the presence of luck using FDR, they were able to estimate how many funds deliver zero, negative, and positive alphas. Barras, Scaillet, and Wermers concluded that around 76.6% of all funds have zero alphas, 21.3% have negative alphas, and only 2.1% have positive alphas. They also argued that the persistence of mutual fund performance is partially driven by persistent expense ratios.

Fama and French (2010) investigated the luck versus managerial skill question by applying the bootstrapping methodology on the mutual funds in the CRSP database between 1984 and 2006. They found that mutual funds underperform the CAPM, 3-factor, and 4-factor models by approximately the expense costs. The funds with weak stock-selection abilities cancel out the funds with superior managerial skills when evaluated on the aggregate level. To investigate individual funds' performance, Fama and French compared the distribution of the actual and bootstrapped t-statistic of alpha. The results reveal that when the returns are evaluated before expense costs, there is an indication of managerial skill for both out- and underperforming funds; however, only a few funds have enough skills to cover the costs when the returns are measured net of fees.

There is an abundance of studies on performance persistence and its economic significance on the US mutual funds. However, scientific literature on the performance of Norwegian mutual funds is lacking. One of the most comprehensive studies using Norwegian data free of survivorship-bias was done by Sørensen (2009). He examined Norwegian mutual funds' performance and performance persistence between 1982 and 2008. He found no statistically significant abnormal returns over the returns obtained with Fama and French model. The bootstrapped results show only weak evidence of skill for the best performing funds and more substantial evidence of bad managerial skill for the worst performers. There is also no evidence for persistence among winners and losers.

Despite the lack of academic literature on Norwegian mutual funds' performance persistence, there is adequate literature written to conduct proper performance evaluation and persistence study on the Norwegian stock market. The study can be of particular interest to investors in Norwegian mutual funds who are paying substantial fees to managers claiming to deliver persistent returns in excess of the passive benchmarks.

3. Data

3.1 Norwegian Mutual Funds

The data on mutual funds are obtained from the Morningstar Direct database. The final sample includes 99 open-end equity funds that have ever existed from 1996 to 2019 and invested at least 80 percent in the Norwegian stock market. The restriction is required to examine the performance of Norwegian funds exclusively. The sample excludes index funds since actively managed funds are the ones that we are interested in investigating. The actively managed funds pursue an active investment strategy and claim to deliver returns above the specified market benchmark.

The number of monthly returns for the selected mutual funds varies from 13 to 288 observations. The monthly return on Morningstar is constructed using the following formula:

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

In which NAV is the monthly net asset value of the fund. All income and capital gain distributions during the period are assumed to be reinvested, while management, administrative, distribution, and other costs are deducted (Morningstar, 2020).

3.2 Risk Factor Loadings

We extract the Norwegian stock market's risk pricing factors from Bernt Arne Ødergaard's website to run Fama and French (1996) and Carhart (1997) benchmark models. Table 1 shows which particular factors are used in the study and their descriptive statistics.

Table 1: Descriptive Statistics of Risk Factors

Both tables display statistics on the risk factors constructed by Bernt Ødegaard that are used in the regression analysis. Panel A shows the average monthly values of the factors in different time periods. The numbers in brackets are p values against the null hypothesis of observations being equal to zero. Panel B shows correlation statistics among the factors throughout the whole time period. OSEFX index is Rm, which is the proxy for market return. The choice of OSEFX as the market return is discussed in section 3.4. The returns are reported in percent.

Panel A: Average values

	1996-2019	1996-2003	2004-2011	2012-2019
SMB	0.62 (0.01)	1.13 (0.00)	0.30 (0.50)	0.44 (0.14)
HML	-0.12 (0.64)	0.05 (0.92)	-0.17 (0.66)	-0.24 (0.49)
PR1YR	1.04 (0.00)	0.45 (0.41)	1.04 (0.02)	1.63 (0.00)

Panel B: Correlations

	Rm	SMB	HML	PR1YR
Rm	1			
SMB	-0.47	1		
HML	-0.2	-0.05	1	
PR1YR	-0.22	0.15	-0.04	1

3.3 Risk-free rate

As a proxy for the risk-free rate, we use the Norwegian Interbank Offered Rate (NIBOR)[1]. NIBOR reflects the interest rate required by lenders for unsecured money within two days of delivery. The rate is calculated as a simple average of submitted interest rates by NIBOR panel banks for each maturity (2020). Historical NIBOR rates are taken from the Norges Bank website for the period from 1995 to 2013, and from the Oslo Børs website for the period 2014-2019 since Norges Bank and Oslo Børs were the official authorities responsible for calculating the money market interest rate. To transform the NIBOR rate to a monthly risk-free interest rate, we apply the following formula:

$$(1 + NIBOR)^{1/12} - 1$$

[1] Bernt Ødegaard also used NIBOR as a proxy for the interest rate in his studies of Norwegian stock market.

3.4 Market Return

The market return can be typically obtained by getting the return of a value-weighted portfolio of all listed stocks. Although in case of the Norwegian stock market, the value-weighted portfolio might be formed only by a handful of big companies. For instance, Telenor, Statoil, and Norsk Hydro constituted 53% of the total stock market in 2006 (Næs et al., 2009). Hence, a choice of performance benchmark as a market portfolio proxy is crucial with the Norwegian stock market because depending on the choice, mutual funds' performance may vary drastically.

Oslo Børs created a capped version of the benchmark investible index - OSEFX. The uncapped version of the index, OSEBX, represents all the shares listed on the Oslo stock exchange. OSEBX exhibits the problem mentioned above – some stocks can skew the overall performance of the index, which makes the index performance less representative of all stocks listed on the stock exchange. As shown in Table 2, OSEFX and OSEBX are highly correlated since OSEFX is constructed based on the OSEBX index. What is noticeable is that OSEFX is highly correlated with the value-weighted index. OSEFX index is like a version of the VW index that is constructed to comply with the restrictions for regulating investments in Norwegian mutual funds. We use OSEFX as the market proxy in our study since it is the most fitting index approximating the market return in Norway. OSEFX allows a maximum weight of the security to be 10% of the total market value of the index, and securities that exceed 5% cannot exceed 40% combined. The monthly returns of the OSEFX index for the period 1996-2019 are obtained from the Oslo Børs website (Oslo Børs, 2020).

Table 2: Descriptive Statistics of Market Indexes

The table illustrates statistics on different market indexes of the Oslo stock exchange. Panel A shows monthly returns in percentage terms for the overall and split time periods. Mean is the average value of monthly returns. St.d is the standard deviation in the sample of monthly returns. Min is the minimum value, while max represents the maximum value in the sample of monthly returns. Med denotes the median value in the sample of monthly returns. EW and VW are equal-weighted and value-weighted indexes constructed by Ødegaard using Norwegian market data. EW and VW indexes are presented in the table to compare the constructed indexes with OSEBX and OSEFX indexes. Panel B shows correlations among the Norwegian market indexes.

Panel A: Monthly returns of indexes

Time period	Index	mean	st.d	min	med	max
1996-2019	EW	1.33	10.07	-18.33	1.43	12.29
	VW	1.66	12.26	-21.04	1.70	16.71
	OSEBX	0.90	13.49	-25.22	1.12	15.83
	OSEFX	0.91	14.35	-27.17	1.13	16.52
1996-2003	EW	1.53	10.27	-18.33	2.12	12.29
	VW	1.94	12.18	-20.55	2.31	16.71
	OSEBX	0.78	13.10	-25.14	1.19	14.02
	OSEFX	0.82	13.28	-25.42	1.30	14.24
2004-2011	EW	1.36	9.37	-16.23	2.02	12.19
	VW	1.79	11.95	-21.04	2.51	14.74
	OSEBX	1.11	13.67	-25.22	1.82	15.83
	OSEFX	1.04	14.53	-27.17	1.64	16.52
2012-2019	EW	1.11	4.32	-6.83	0.85	6.31
	VW	1.24	5.23	-7.90	1.51	8.26
	OSEBX	0.81	5.47	-8.75	1.00	8.18
	OSEFX	0.85	5.41	-8.83	0.72	7.86

Panel B: Correlations

	EW	VW	OSEBX	OSEFX
EW	1			
VW	0.9	1		
OSEBX	0.89	0.98	1	
OSEFX	0.9	0.97	0.99	1

3.5 Survivorship Bias

The sample data includes all the funds that have ever existed for at least 12 months, from 1996 to 2019. The inclusion of dead funds is crucial to conduct an accurate analysis of the mutual funds' performance in the specified time period. If a sample has only survived mutual funds, that sample's overall performance will be positively skewed and not entirely representative of the reality (Brown, Goetzmann, Ibbotson & Ross, 1992). Additionally, the studies conducted on the samples that exclude dead funds indicate predictability in the funds' returns (Brown et al., 1992; Carpenter & Lynch, 1999). Those mutual funds indicate performance persistence mainly due to survivorship bias rather than managerial ability to generate excess risk-adjusted returns (Malkiel, 1995). Therefore, we use survivorship bias-free data to avoid adverse effects when dead funds are excluded.

4. Methodology

4.1 Measuring Performance

One of the main questions in mutual fund literature is whether the active mutual fund managers can consistently beat the market and add value for investors. The initial step to address this question is to find a reliable proxy for measuring mutual funds' performance.

The historical raw returns are widely used to measure the funds' performance. Carhart, who also used raw returns in his study, argues that although risk-adjusted returns are more likely to assess stock selection skills more accurately, the persistence estimation from the asset pricing model is exposed to model biases (Carhart, 1997).

Despite concerns regarding the model biases, assessing mutual fund performance based on raw returns does not consider the level of risk taken by funds since high raw returns might be associated with higher risk. Therefore, the analysis based on raw returns can mislead investors to prefer the high-risk mutual funds over the low-risk ones to obtain higher returns while disregarding the investment riskiness. Blake and Timmermann (2003) explain the importance of using risk-adjusted returns in evaluating fund performance, performance persistence, and recognizing managers’ stock-picking ability. They argue that measuring performance without adjusting for risks may result in ranking funds based on their systematic risk level rather than the level of managers' stock selection ability. That is not in the best interest of risk-averse investors since riskier funds also have a high probability of underperforming. Moreover, replicating and implementing high-risk strategies does not require managerial skill, thus using raw returns is less likely to contribute to identifying funds’ manager superior skills. Blake and Timmermann suggest that it is essential to provide investors with reliable performance analysis based on risk-adjusted returns to identify persistent underperformers and less-constant outperformers. It allows them to revise their investment strategies accordingly (Blake & Timmermann, 2003). Hence, it is crucial to assess fund performance based on risk-adjusted returns.

We utilize both raw and risk-adjusted returns to attain a more precise estimation of the Norwegian mutual funds' performance and mitigate the biases that arise due to models' selection for measuring performance.

4.1.1 Risk-Adjusted Return

Jensen's single-factor model (1968) is developed based on Sharpe's Capital Asset Pricing Model (CAPM) that explains the relationship between risk and expected-returns for a given asset. In the Jensen model, the return in excess of the risk-free rate is adjusted for the market risk by including market risk exposure. The intercept of the model represents an abnormal return. A significant positive alpha implies the managers' ability to forecast equity prices, meaning that a captured return is higher than the expected return adjusted for risk exposure. In contrast, a negative alpha denotes mutual fund underperforming and perverse forecasting ability of fund managers.

Even though the single-factor model defines an exact data generating process to describe the excess return produced by mutual funds managers, it only reflects market risk factors. Several scholars claim that Jensen's single-factor model does not measure the mutual funds' performance accurately since it only takes market proxy as the risk factor into account and it is not able to capture cross-sectional differences in average stock returns accurately (Elton et al.,1993; Fama & French, 1993; Carhart, 1997). Therefore, the single-factor model is not an appropriate benchmark to assess mutual funds' performance in producing an excess return.

4.1.2 Fama and French 3-factor Model

Multi-factor models are developed to address the problem that arose from the empirical evidence that a single market risk factor cannot fully explain expected returns. In light of that evidence, Fama and French examined Jensen's model using the US stock market data. They suggest that the market risk is not the only relevant risk factor to explain the returns, and other systematic risk factors also affect stock performance. Fama and French created their multi-factor model by taking Jensen's model and adding two non-market risk factors to it; size factor (the higher average return of small-cap stocks relative to large-cap stocks) and book to market value factor (the higher average return of value stocks relative to growth stocks) (Fama & French, 1993).

Fama and French argue that a 3-factor model can enhance the model's specification and ability to describe the typical variation in stock returns considerably compared to the single-factor model. Fama and French’s study indicates a robust negative relationship between stock returns and size - smaller firms are more likely to have higher average returns. It also reveals a substantial positive correlation between stock returns and book-to-market value ratio (Fama & French, 1992, 1993).

4.1.3 Carhart 4-factor Model

Carhart argues that the 3-factor model is unable to explain cross-sectional variations in momentum-sorted portfolio returns. He developed a 4-factor model by using the Fama and French 3-factor model and adding a one-year momentum factor (PR1YR), which previously was introduced by Jagadeesh and Titman (1993). PR1YR defines the difference between the average returns of the past best-performing and worst-performing portfolios. Carhart's model reveals whether fund managers possess the stock-picking ability or outperform the market by merely following a zero-investment strategy of investing in a portfolio that would long stocks with the highest past one-year return and short stocks with the lowest past one-year return. The PR1YR factor indicates that the stocks that performed well in the prior year are expected to achieve higher returns relative to those that performed poorly (i.e., contrarian stocks) (Carhart, 1997).

Carhart argues that the 4-factor model is more efficient in assessing the active mutual funds' performance against passive funds' performance since it captures the average return obtained from fundamental trading strategies that bet on stock’s beta, market capitalization size, book to market, and momentum factors. Therefore, Carhart's model serves as the most reliable benchmark to represent the passive funds' performance and deliver more precise alphas to evaluate active fund managers' stock selection ability (Carhart, 1997).

We employ the Carhart 4-factor model as the primary model and the Fama and French 3-factor as the alternative model to measure the mutual funds' performance. Then, we discuss the ability of these two models to evaluate performance of actively managed Norwegian mutual funds.

4.2 Bootstrapping Method

The performance that has been measured by the asset pricing models does not reflect the role of luck in performance outcomes. It is expected that some funds beat the passive benchmark simply by chance or underperform the passive indices as a result of bad luck. The statistical inference of excess return obtained from the factor benchmark models is valid under OLS model assumptions. The validity is questioned with mutual funds' returns since the assumption of the normal distribution of residuals is violated (Kosowski et al., 2006). The combination of non-normality in both time-series of individual mutual fund returns and cross-section of mutual funds alpha makes t-statistics invalid to test the hypothesis of the existence of abnormal performance.

For individual funds, non-normality in the distribution of residuals arises from skewness and kurtosis in the distribution of returns, time-series autocorrelations in returns, and dynamic investment strategies that lead to fluctuation in funds' risk-taking levels over time. The non-normality of individual fund alphas leads to nonnormality in the cross-section of fund alphas. Besides, the uneven distribution of risk-taking level among mutual funds affects the cross-sectional distribution of alphas. When a large group of high-risk funds is present in the sample, the cross-sectional distribution of alphas is thicker than normal distribution in the tails, leading to over rejection of the null hypothesis of no true excess return. In contrast, when a large group of low-risk funds exists in the sample, the cross-sectional distribution of alphas is thinner in the tail resulting in under-rejection of the null hypothesis of no excess return (Kosowski et al., 2006).

To differentiate between luck and superior managerial skills in generating excess returns, we employ a bootstrap method that estimates the sample distribution of t-statistic without making assumptions about the underlying population. This method infers about the population parameter based on the sample statistic by multiple random resampling from the original sample data. The bootstrapping method provides a randomly constructed baseline of performance representing excess returns produced only by luck. This baseline is created by simulating the returns assuming that true alpha is equal to zero. By comparing the distribution of bootstrapped and actual estimates, we can answer whether over or underperforming is because of luck or managers' skill. If bootstrap iterations create far fewer extreme positive values of alphas or alphas t-statistics than the actual ones, it implies that luck is not the only source of high abnormal returns, and managers possess superior stock-selection skills to outperform the passive indices.

We employ the bootstrapping method developed by Fama and French (2010) based on the initial bootstrapping method of Kosowski et al. (2006). The Kosowski method generates independent simulations for individual funds, while the Fama and French method run the simulations by jointly resampling the factor returns and residuals for all funds. This resampling method controls for the effects of correlated movement in 4-factor explanatory returns volatilities and residuals in the model. This modification captures any correlation across estimated alphas arising from a benchmark model that does not capture all common variation in fund returns. (Fama & French, 2010).

The findings of these methods indicate variation in the skill across actively managed mutual funds. Fama and French (2010) found no evidence of managerial skill, while Kosowski et al. (2006) detected a small number of skilled managers. Blake et al. (2017) compared the two approaches using the same dataset and survival rule to identify whether the different results come from the different research periods, distinct inclusion criteria, or the bootstrap methodology. By employing Kosowski et al. (2006) method, they detected a small number of funds with superior manager skills. However, they discovered a little indication of outperformance when jointly resampling the fund and factor returns, as suggested by Fama and French (2010).

This comparison suggests that employing the Fama and French bootstrapping model controls for the systematic relationship between the funds' returns and the factor benchmarks in addition to non-systematic risk contained in the benchmark models' residuals. Therefore, by employing this method, we set marginally stricter measures when classifying funds' managers as "stars" with superior skills. (Blake et al., 2017).

4.2.1 Fama and French Bootstrap Implementation

We start by running the Carhart 4-factor model using individual funds' monthly returns across the study period and saving the actual estimated alphas, t-statistics of alphas, risk factors' coefficients, and residuals.

In the following step, a ($T \times 1$)-dimension vector of a random sample of the monthly returns data point is drawn from the uniform distribution $U_t(0,1)$ that produce values between 0 and 1. This vector is multiplied by T, and its components round up to the nearest integer. This constructed matrix represents the ordering of monthly returns in the sample. T denotes the number of periods in our sample (T=288). It generates a matrix that represents a vector of time indices drown randomly and with replacement from the time points in the data set:

$$T_b = round(T \times \{U_t(0,1)\}_{t=1}^{T}) \qquad b = 1, \dots, 10000$$

In each simulation irritation, a new series of risk factors $F(T_b)$ with the dimension of (T×K) is created according to the drawn time indices, where K is the number of risk factors in the model. The same procedure is repeated for the saved residuals obtained from the primary regression model by building a matrix of (T×N), where N is the total number of funds. Since the number of monthly return observations varies across funds and there some missing data in the sample, if the randomly draws produce less than 12 observations for a fund, it is not included in that bootstrap iteration.

We generate a new series of funds' returns with the null hypothesis of zero alpha using the original estimates of coefficients for the risk factors along with the bootstrapped factor returns and residuals from simulations. The constructed funds' returns are regressed against the corresponding risk factors and generate 10000 alpha estimations and t-statistics of the alpha for each of the 99 funds. The average of the bootstrapped results represents how the funds perform (net of fees) when there is no managerial skill, and excess returns are created only by luck.

Lastly, we compare the distribution of t(α) estimates from funds' returns with the distribution from bootstrap simulations to examine if luck or managers' skill generates the excess returns. We compare the average of all simulated t(α) and the actual estimates at the different percentiles to analyze the bootstrapped outcomes at various levels of funds' performance- from the worst-performing to the best performing funds.

We also calculate the bootstrapped (p-value) of the estimations, which is the fraction of 10000 simulations that produce alphas or t-statistic of the alphas greater than the actual estimates at different percentiles. We test the null hypothesis that true alpha is equal to zero using the bootstrapped p-values at 5% significance level. It means that for the right tail quintiles if less than 5% of the simulations produce t(α) above the actual t(α) estimate, we can conclude that some managers possess sufficient skills to attain positive excess returns. In contrast, if this fraction in the left tail quintiles exceeds 95%, we can argue that managers possess inferior skills to obtain excess returns above the passive benchmarks.

4.3 Persistence Tests

Research on performance persistence in mutual funds provides an essential methodology for studying the ability of actively managed funds to outperform the market and consistently attract new investors. Despite the extensive literature on mutual funds ‘performance persistence, previous studies demonstrate contradictory findings regarding the existence of persistence in actively managed mutual funds.

We test for performance persistence using three methodologies that are frequently found in the literature: recursive portfolio test, contingency table, and cross-product ratio.

4.3.1 Recursive Portfolio Formation Method

We primarily assess the persistence in mutual fund performance by employing a recursive portfolio formation approach. This method has been used in several leading mutual performance studies written by Hendricks et al. (1993), Grinblatt and Titman (1993), and Carhart (1997). Carpenter and Lynch examine the specification of different persistence tests in the absence of survivorship bias. They suggest that the recursive portfolio formation method is well-specified in measuring performance persistence, particularly for the top- and bottom-ranked portfolios (Carpenter & Lynch, 1999).

This method includes sorting the funds in quintiles based on their performance in the ranking period, forming equally weighted (EW) portfolios of mutual funds in each quintile, and then evaluating the performance of formed portfolios over the consequent holding period before a new portfolio is created based on the similar process.

We measure the performance of the mutual funds based on risk-adjusted returns using the Carhart 4-factor model. The excess return is considered as a mutual fund's manager's ability to outperform the Carhart 4-factor benchmark. To obtain more reliable estimations from the model, we set a minimum 24 observations requirement for the ranking period and then extend it to 36 observations to examine how a longer ranking period will affect the persistence measurement. We rank funds based on both alphas and t-statistics of alphas obtained from the Carhart 4-factor model over the ranking period.

We evaluate persistence for short and long time horizons by holding the constructed portfolios in each quintile in the subsequent 3 to 24 months. At the end of each holding period, we rebalance the portfolios according to lagged 24-months to 36-months alpha and t-statistics and repeat it for the whole study period. We follow the Carhart approach (1997) to control for the survivorship and look ahead bias by including the funds existing in ranking periods and terminating during holding periods. The EW returns of portfolios are readjusted for the funds that disappear based on the remaining funds in each quintile. Lastly, we examine performance persistence of mutual funds across each quintile by employing the Carhart 4-factor model for the time-series of portfolios' EW returns over the holding periods. We study the alphas to see if the best-performing quintile portfolios persistently outperform the benchmark and if the worst-performing quintile portfolios constantly underperform the benchmark model. We also examine statistical significance of an estimated alpha for quintile portfolios by using the bootstrapped p-value. The bootstrapped analysis gives an insight into whether quintile portfolios' alphas are generated by luck or managerial skills. The significant quintile portfolios' risk-adjusted returns imply that persistence or inconsistent performance is because of the fund managers' actions.

4.3.2 Contingency Table

We test the consistency in mutual fund ranking by constructing contingency tables of initial and subsequent 12-month, 24-month, and 36-month mutual fund rankings by following the Carhart approach (1997). We rank funds based on the annual raw returns net of fees for the 12-month period and based on the Carhart 4-factor model's alphas for 24-month and 36-month periods.

It is worth mentioning that ranking on raw returns may reflect the persistent systematic risk taken by fund managers while ranking on risk-adjusted returns differentiates managerial skill more precisely and picks up the model bias between ranking and subsequent periods. That is why we base our ranking both on raw and risk-adjusted returns. We expand the time-window from 12 to 36 months since the regression coefficients will be more meaningful with a longer time horizon and also because one-year persistence can be a noisy measure (Carhart, 1997).

After ranking, we place the funds in quintile over an initial and subsequent time window. Then, we count the number of times a fund is observed in one of the quintiles in the subsequent period, conditional on its ranking over the initial period. Funds that disappear during the subsequent period are placed in a separate category for dead funds.

The bars in the contingency table for initial rank i and subsequent rank represent the conditional probability of achieving a subsequent ranking of quintile j (or dying) given its initial ranking of i.

4.3.3 Cross-Product Ratio

The contingency tables in the previous section illustrate the mutual funds' performance persistence visually but do not demonstrate statistical significance of the results. Therefore, we employ the non-parametric test proposed by Brown and Goetzmann (1995), which applies a cross-product ratio (CPR) test to evaluate persistence in the sample.

By following this approach, we categorize the performance of mutual funds based on 12-month net returns and based on the risk-adjusted 24-month and 36-month returns using the Carhart 4-factor model. The funds evaluated on raw net returns are categorized as winners and losers in the following way - a fund is a winner if its net return is greater than the median net return of all the funds in that period; otherwise, it is a loser. The funds assessed by the risk-adjusted returns are categorized as winners if the fund's alpha is greater than the median risk-adjusted return of all the funds in that period; otherwise, it is a loser. After categorizing the funds, we construct two-way contingency tables for each time interval that indicate the mutual fund's performance in two consecutive time periods. The funds are labeled as WW (LL) if they are winners (losers) over an initial and the following periods. If a fund is categorized as a winner (loser) in the first period and as a loser (winner) in the second period, the fund is defined WL (LW). The repeat performers imply funds performing persistently from the prior interval to the following interval, while reverse performers do not continuously remain in the same category at different intervals (Brown & Goetzmann, 1995).

We aggregate the resulting contingency tables and conduct the following test on the aggregated absolute frequencies denoted by N_{WW}, N_{WL}, N_{LW} and N_{LL} where the sum of the absolute frequencies is $N = N_{WW} + N_{WL} + N_{LW} + N_{LL}$.

The cross-product ratio captures the odds ratio of the funds that persist in their performance to the ones that switch their performance in the subsequent period. The cross-product ratio (CPR) is calculated as follows:

$$CPR = \frac{N_{WW} \times N_{LL}}{N_{WL} \times N_{LW}}$$

CPR is greater than one if the number of funds with the same performance in two consecutive periods is higher than the number of reverse performers. CPR being equal to one implies persistence in mutual funds' performance does not exist since, in the absence of persistence, four categories denoted by N_{WW}, N_{WL}, N_{LW} and N_{LL} would have 25% of the total number of funds (Goetzmann & Ibbotson, 1994).

We test the statistical significance of the CPR being equal to one under the null hypothesis representing the absence of persistence. Since the CPR is assumed to have the natural log form, Z-statistic is computed as below, using a natural log odds ratio and standard error of the natural logarithm of the CPR (Christensen, 1990).

$$\sigma_{\ln(CPR)} = \sqrt{\frac{1}{N_{WW}} + \frac{1}{N_{WL}} + \frac{1}{N_{LW}} + \frac{1}{N_{LL}}}$$

$$Z = \frac{\ln(CPR)}{\sigma_{\ln(CPR)}}$$

Thus, if the Z-statistic is greater than the critical value at 5% significance level, the null hypothesis is rejected in favor of the presence of persistence.

To examine the robustness of the results, we also conduct a chi-square test (Agarwal & Naik, 2000) that examines the funds' performance persistence by comparing the observed frequency distribution of N_{WW}, N_{WL}, N_{LW} and N_{LL} with the expected frequency distribution.

We obtain the chi-square statistic in the following way:

$$\chi^2 = \frac{(N_{WW} - D1)^2}{D1} + \frac{(N_{WL} - D2)^2}{D2} + \frac{(N_{LW} - D3)^2}{D3} + \frac{(N_{LL} - D4)^2}{D4}$$

Where

$$D1 = \frac{(N_{WW} + N_{WL}) \times (N_{WW} + N_{LW})}{N}, D2 = \frac{(N_{WW} + N_{WL}) \times (N_{WL} + N_{LL})}{N},$$

$$D3 = \frac{(N_{LW} + N_{LL}) \times (N_{WW} + N_{LW})}{N}, \quad D4 = \frac{(N_{LW} + N_{LL}) \times (N_{WL} + N_{LL})}{N}$$

5. Results

5.1 Measuring performance

We start presenting the results in Table 3 by showing the performance of the EW portfolio of all fund returns, which is an overview of how mutual funds perform in aggregate when measured using OSEFX as a market proxy and regressed using CAPM, Fama and French 3-factor, and Carhart 4-factor benchmarks.

The market beta is statistically significant and close to one as expected since the sample represents the Norwegian market's return as we imposed the restriction of investing solely in Norwegian equities. SMB factor is statistically significant and positive in all models, which means that the EW portfolio's return can be partially explained by exposure to the SMB risk factor. SMB factor is positive, indicating that on average, funds bet on small over big capitalization equities. HML factor is negative, demonstrating that, on average, funds prefer low book-to-market over high book-to-market stocks. Nevertheless, since the HML factor is close to zero and significant only at 10% in Fama and French model and insignificant in the Carhart model, the funds have little to no exposure to the risk factor. PR1YR factor is insignificant; thus, the EW portfolio of mutual fund returns cannot be explained by the momentum factor. The regression models used have high explanatory power, which is fairly similar across the models since the Fama and French and Carhart models have the same adjusted R-squared. At the aggregate level, market, SMB, and to a small extent, HML risk factors explain the mutual funds' returns. All of the alphas obtained are insignificant, meaning that the EW portfolio of mutual funds fails to deliver the returns in excess of those explained by the benchmarks.

Our aggregate results are aligned with the findings of Bernt Ødergaard (2009), who discovered that HML and momentum factors are irrelevant in the Norwegian stock market. Similarly, Fama and French (2010) showed that the US mutual funds have little exposure to HML and momentum factors, in aggregate. Additionally, our finding that the aggregate portfolio of mutual funds generates no excess returns for investors is similar to the findings made in the previous studies on the US market, for instance, in Jensen (1968), Malkiel (1995), and Fama and French (2010).

Table 3: Aggregate Fund Performance

The table displays alphas, market exposure coefficients, risk factor coefficients, and the adjusted R-squared values obtained from the regressions on aggregate EW portfolio of excess monthly fund returns. The values in brackets are corresponding t-statistics that were adjusted with the Newey-West procedure to account for autocorrelation and heteroskedasticity (Newey & West, 1987). The alphas are in percent per year. The sample period is from 1996 to 2019.

Model	α	β_{MKT}	β_{SMB}	β_{HML}	β_{PR1YR}	Adj.R^2
CAPM	0.89	0.94***				0.97
	(1.25)	(55.52)				
Fama&French	-0.60	0.98***	0.15***	-0.02*		0.98
	(-1.01)	(70.35)	(10.38)	(-1.68)		
Carhart	-0.69	0.98***	0.15***	-0.02	0.01	0.98
	(-1.15)	(71.04)	(10.81)	(-1.53)	(0.50)	

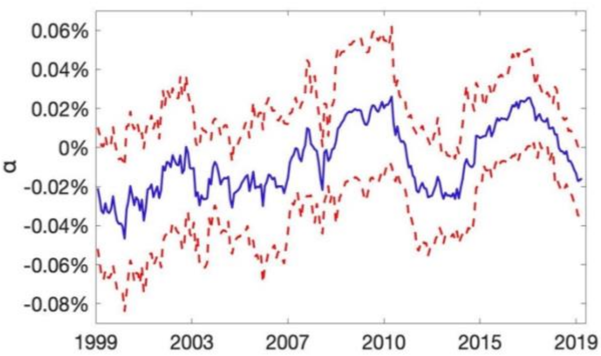
Significance codes: p=1% (***), p=5% (**), p=10% (*)

To examine Norwegian mutual funds' average performance over time, we illustrate rolling and extending alpha estimation with a window of 36-month returns in Figure 1. The upper panel reports the alpha obtained from the Fama and French 3-factor model, while the lower panel shows the alpha estimation from the Carhart 4-factor model. The left graph in each panel indicates alphas for the rolling window, whereas the right graph represents alphas for the extending window. From Panel A1, the 3-factor estimations of alpha in the rolling window had been growing from -0.02 to 0.02%. The rolling window alpha estimations in Panel B1 display a similar trend; however, the 4-factor model's alphas are smaller than the alphas of the 3-factor model during the rise and fall from 2015 to 2018. It infers that the difference in alphas can be partially explained by the momentum factor captured by the Carhart model throughout this period.

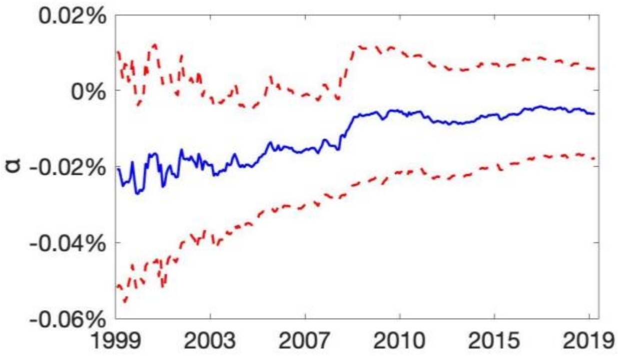
Looking at Panel A2 and Panel B2, we observe that extending window alpha estimates at an aggregate level from the 3-factor and 4-factor models are identical and below zero over time. The similar alphas of both models indicate that the PR1YR risk factor is not significant at the aggregate level.

Figure 1: Rolling and Extending Window Alphas of the Equally Weighted Portfolio

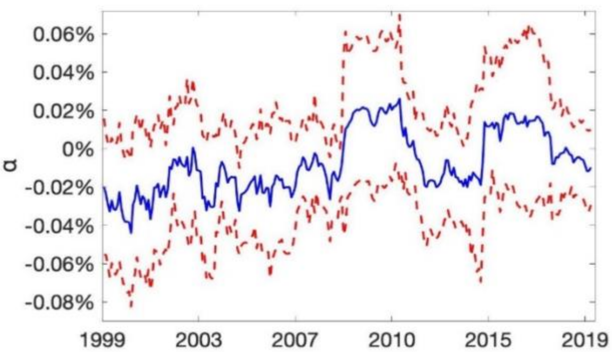
Panel A1: Rolling window 3-factor alphas



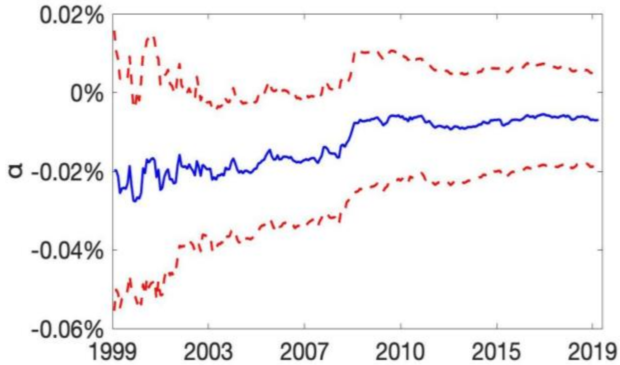
Panel A2: Extending window 3-factor alphas



Panel B1: Rolling window 4-factor alphas



Panel B2: Extending window 4-factor alphas



This figure displays alpha estimates of the equal-weighted portfolio of all mutual fund returns with the window width of 36 months. Panel A reports the alphas of the Fama and French 3-factor model, and Panel B reports the alphas of the Carhart 4-factor model. In each panel, the left graph represents rolling window alpha estimates, and the right graph represents extending window alpha estimates. The solid line indicates the alpha estimates, while the dotted line is the Newey-West-corrected two standard error bands. The sample period is 1996 to 2019. The alpha estimates are annual and in percent.

5.2 Luck versus skill in the mutual fund returns

Our finding that the Norwegian mutual funds do not deliver excess returns above the passive benchmarks at an aggregate level might not be robust. Additionally, some individual funds may deliver abnormal returns above the specified risk factors. To test robustness of the obtained results and establish if any funds deliver abnormal returns, we employ the Fama and French's (2010) bootstrap simulation method. The procedure aims to detect any abnormal returns and whether those returns are the consequence of managerial skill or simply luck. The results of the bootstrap procedure are presented in Table 4. Fama and French advise using simulation on the t-statistics rather than alpha estimates due to differences in the number of observations and residual variance. Since the t-statistic of alpha reflects precision with which alpha is obtained, it is viewed as a more reliable estimate. We are following the researchers' recommendations but present the results based both on alphas and t-statistics.

The results show that the simulated alphas of the four worst performers and extreme percentiles of the worst-performing funds are above the actual estimates in at least 90% of the cases. The result indicates underperformance of the worst-performing mutual funds, which is due to lack of skill rather than luck. The rest of percentile portfolios and the best performing funds have insignificant alphas; thus, we cannot reject the null hypothesis, and the true value of the alphas is zero. Nonetheless, as we have established earlier, the simulation results on t-statistics are more reliable.

The simulation on t-statistics presents the results that are more promising for the worst-performing funds but are also less favorable for the best performers. The worst performing funds show no significant results since the number of simulated t-statistics above the actual estimate is mostly below 90%. The percent of simulated values above the actual estimate among the best performers increased compared to the value with alpha estimates, which indicates even less evidence of skill for the best performers. Nonetheless, all of the percentile portfolios, along with the worst and best funds, show insignificant performance at 5% significance level. Therefore, we cannot reject the null hypothesis of zero alpha for those funds.

What is important to note is that we are using returns net of fees, which we obtained from the Morningstar database. Thus, we cannot say if managers have sufficient skills and can produce excess returns before charging a management fee. The bootstrap analysis displays evidence of weak managerial skill based on bootstrapped alpha and no skill according to bootstrapped t-statistic among the Norwegian fund managers.

Table 4: Bootstrap Analysis of the Mutual Funds Returns

The table presents the actual and bootstrap simulated alphas produced with the Carhart 4-factor model along with the actual and bootstrap simulated t-statistics of the Norwegian mutual funds. The first column shows for which funds the estimates are presented - five worst-performing funds, 1 to 99% deciles of the funds, and five best-performing funds. Simulated column displays the average values across 10000 simulations for the specified fund or percentile. Sim>Act column illustrates how many of the mutual funds' simulated estimates are above the actual estimates. The bold estimates in the column indicate 5% significance level. The t-statistics were adjusted with the Newey-West procedure to account for autocorrelation and heteroskedasticity (Newey & West, 1987). The alphas are monthly and in percent. The sample period is from 1996 to 2019.

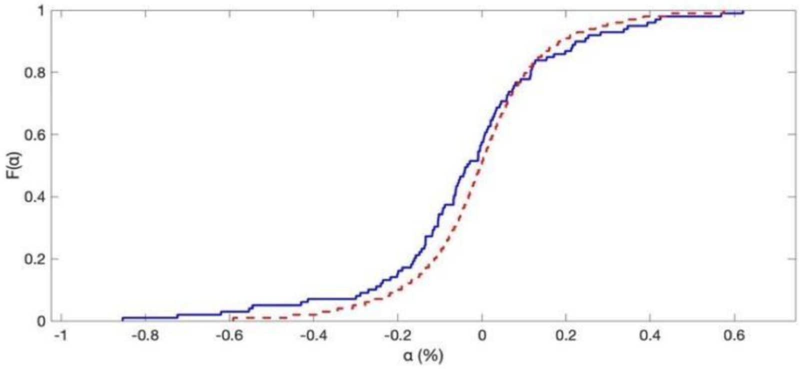
	Alphas			t-statistics		
	Actual	Simulated	Sim>Act	Actual	Simulated	Sim>Act
Worst	-0,87%	-0,59%	88,50%	-3,23	-2,77	77,32%
2nd	-0,70%	-0,45%	91,78%	-2,81	-2,24	82,27%
3rd	-0,63%	-0,39%	93,13%	-2,76	-2,01	89,74%
4th	-0,55%	-0,34%	94,02%	-2,45	-1,84	86,38%
5th	-0,55%	-0,30%	**96,72%**	-2,41	-1,72	90,01%
1%	-0,79%	-0,53%	90,37%	-3,02	-2,52	79,24%
2%	-0,67%	-0,43%	92,54%	-2,79	-2,14	85,78%
3%	-0,59%	-0,37%	93,77%	-2,61	-1,94	88,16%
4%	-0,55%	-0,33%	**95,38%**	-2,43	-1,79	87,96%
5%	-0,49%	-0,29%	**95,27%**	-2,38	-1,68	90,46%
10%	-0,26%	-0,20%	79,90%	-1,92	-1,29	90,58%
20%	-0,16%	-0,11%	79,66%	-1,18	-0,85	77,76%
30%	-0,11%	-0,07%	81,45%	-0,89	-0,53	79,86%
40%	-0,07%	-0,03%	75,97%	-0,55	-0,27	75,23%
50%	-0,03%	0,00%	74,40%	-0,25	-0,03	70,47%
60%	0,01%	0,03%	64,85%	0,07	0,22	64,00%
70%	0,04%	0,06%	59,88%	0,40	0,49	57,11%
80%	0,12%	0,10%	40,76%	0,86	0,80	43,65%
90%	0,24%	0,19%	25,28%	1,29	1,25	44,47%
95%	0,37%	0,28%	19,27%	1,70	1,64	42,85%
96%	0,40%	0,32%	21,22%	1,87	1,76	39,23%
97%	0,42%	0,36%	29,45%	2,06	1,92	36,61%
98%	0,50%	0,41%	24,72%	2,26	2,11	36,34%
99%	0,60%	0,51%	27,89%	2,32	2,46	50,87%
5th	0,39%	0,30%	17,87%	1,83	1,68	36,09%
4th	0,41%	0,33%	23,04%	1,90	1,81	40,56%
3rd	0,42%	0,38%	33,17%	2,21	1,99	32,19%
2nd	0,57%	0,44%	19,29%	2,30	2,21	39,65%
Best	0,62%	0,57%	34,33%	2,34	2,70	58,32%

Figure 2 illustrates the Cumulative Density Functions (CDFs) of simulated and actual 4-factor alphas and alpha t-statistics in Panel A and B, respectively. The graph in Panel A indicates that in the left tail of performance, there is less probability mass for the actual distribution than the bootstrap distribution, which means there are weak-performing funds' managers whose actions lead to true negative alphas. However, in the right tail at 80% percentile till 98% percentile there is some indication of superior managerial skills to beat the passive benchmark.

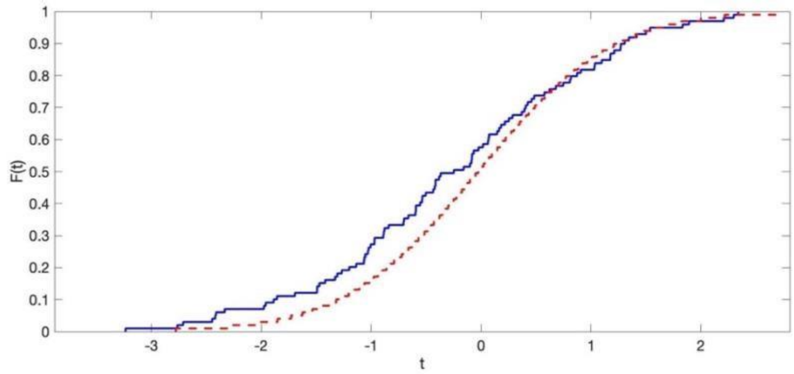
In Panel B, we observe similar results for the left tail that implies true inferior managerial skills. In contrast, in the right tail, the actual t-statistics of alpha estimations are close to simulation results, suggesting that some fund managers possess only adequate skills to generate risk-adjusted returns to cover the fees, but they do not have superior stock-picking abilities to outperform the passive indices.

Figure 2: Empirical Cumulative Density Functions for Simulated and Actual 4-factor Alphas and t-statistics

Panel A: Simulated and Actual CDF of Alphas



Panel B: Simulated and Actual CDF of t-statistics



The figure shows the actual and average simulated CDFs of the 4-factor Carhart model alphas and the corresponding t-statistics for the net returns. The solid line represents the actual values, while the dotted line represents the simulated values. The sample period is from 1996 to 2019.

5.3 Performance Persistence

Findings in the previous section demonstrate that the Norwegian mutual funds generally fail to deliver positive excess returns to investors over the study period. Nevertheless, performance persistence analysis may reveal that a group of funds persistently outperform the benchmark and thus be of interest to investors since they can identify the best performing persistent funds in advance and receive excess returns.

We present performance persistence of the portfolios of mutual funds ranked on lagged 1-year returns in Table 4. The portfolios do not demonstrate substantial variation in the returns, particularly with the Fama and French 3-factor model. Overall, the coefficients of 1-5 spread portfolio are insignificant, and the portfolio has low explanatory power of the returns. In contrast to the spread portfolio with Fama and French, the spread 1-5 Carhart portfolio has higher explanatory power of 11%. Additionally, the market and momentum risk factors are highly significant and explain the variation in the returns between the top and bottom quintile funds.

What is noteworthy is that the momentum risk factor is highly significant for the best and worst-performing quintile portfolios created using Carhart 4-factor model. The best portfolio of funds has positive PR1YR, meaning that the funds follow the momentum strategy of longing top past year performers and shorting worst past year performers. The worst portfolio of funds has negative PR1YR, which means the funds failed the same strategy, possibly due to a reversal of top winners' performance.

The yearly alphas do not differ from zero for all of the portfolios except for portfolio 5 in Fama and French that is being significant at 10% significance level. However, significance disappears after the inclusion of the momentum factor, which partially explains the alpha intercept presented in the 3-factor model. Therefore, there is no indication of yearly performance persistence in our analysis of Norwegian mutual funds. The finding clashes with the one that Carhart got: the alpha of the spread portfolio between the best and worst-performing US mutual funds is highly significant, thus displays one-year persistence in the funds returns.

Table 5: Portfolios of Mutual Funds Grouped on Lagged 1-year Return

The table shows returns, alphas, market exposure coefficients, risk factor coefficients, and the adjusted R-squared values derived from the regressions on 6 portfolios, including the spread portfolio. The alphas are in percent per year. Portfolio 1 is the portfolio of the top-performing funds, portfolio 2 is the portfolio of the next best performing funds, and so on until portfolio 5, which is the worst-performing funds portfolio. 1-5 spread is the hypothetical portfolio of buying top-performing portfolios and shorting worst-performing portfolios aimed to highlight the differences between the two. The table has two panels: one displays the results from Fama and French regression model, and the other shows the results from the Carhart 4-factor regression model. The values in brackets are corresponding t-statistics that were adjusted with the Newey-West procedure to account for autocorrelation and heteroskedasticity (Newey & West, 1987). The sample period is from 1996 to 2019.

Panel A: Fama and French 3-factor model

Portfolio	α	β_{MKT}	β_{SMB}	β_{HML}	Adj.R^2
1	-0.46	1.01***	0.22***	-0.02	0.94
	(-0.47)	(64.35)	(6.80)	(-1.19)	
2	-0.40	0.99***	0.14***	0.01	0.97
	(-0.44)	(62.22)	(5.48)	(0.54)	
3	-0.63	0.98***	0.08***	-0.03***	0.97
	(-0.96)	(64.02)	(5.63)	(-2.10)	
4	-0.78	0.99***	0.13***	-0.03	0.97
	(-1.07)	(49.13)	(5.81)	(-1.47)	
5	-1.79*	0.96***	0.19***	-0.02	0.94
	(-1.95)	(35.57)	(6.06)	(-0.66)	
1-5	1.33	0.04	0.02	-0.03	0.01
spread	(0.95)	(1.40)	(0.45)	(-0.81)	

Significance codes: p=1% (***), p=5% (**), p=10% (*)

Panel B: Carhart 4-factor model

Portfolio	α	β_{MKT}	β_{SMB}	β_{HML}	β_{PR1YR}	Adj.R^2
1	-1.44	1.03***	0.21***	-0.02	0.07***	0.95
	(-1.44)	(70.20)	(7.72)	(-0.87)	(3.27)	
2	-0.67	0.99***	0.14***	0.01	0.02	0.97
	(-0.78)	(64.66)	(5.34)	(0.70)	(1.56)	
3	-0.89	0.98***	0.08***	-0.03*	0.02	0.97
	(-1.41)	(65.30)	(5.52)	(-1.87)	(1.10)	
4	-0.31	0.98***	0.13***	-0.03*	-0.03*	0.97
	(-0.39)	(49.95)	(6.39)	(-1.67)	(-1.82)	
5	-0.96	0.96***	0.20***	0.00	-0.06***	0.94
	(-0.87)	(37.05)	(6.98)	(-0.08)	(-2.96)	
1-5	-0.48	0.06***	0.02	-0.01	0.14***	0.11
spread	(-0.37)	(2.55)	(0.38)	(-0.49)	(5.38)	

Significance codes: p=1% (***), p=5% (**), p=10% (*)

5.4 Recursive Portfolio

In this section, we form recursive portfolios to identify performance persistence in mutual funds. We analyze performance of equally weighted (EW) quintile portfolios created by ranking the funds based on lagged the 24-month and 36-month alpha and corresponding t-statistic of alpha obtained from the Carhart 4-factor model. The EW portfolios are rebalanced after 3, 6, 12, and 24-month holding periods and this procedure is repeated over the whole time period. We examine performance of the EW quintile portfolios to identify how persistent these portfolios are in generating positive or negative alphas. We employ the Carhart 4-factor model to evaluate the risk-adjusted return of portfolios and then apply a bootstrapped method to test whether persistence or inconsistent performance arises due to luck or managers' stock-picking abilities.

Tables 6 reports the results of the recursive portfolio method when funds are ranked and placed in the top to bottom quintiles based on their alphas. Panel A and B show the results based on 36 and 24 ranking periods, respectively, for quintiles 1 to 5, where the first quintile denotes the portfolios containing the best-performing funds, and the fifth quintile includes the worst-performing funds over the ranking period. We also construct 1-5 spread quintile, a hypothetical self-financing quintile[2] that represents the trading strategy of longing the best-performers and shorting the worst-performers. The column estimates are the Carhart 4-factor model alpha and Fama and French (2010) bootstrapped p-value[3] of alpha for 3 to 24-month holding periods. The bootstrapped p-value of alpha indicates if the true risk-adjusted return of quintiles

[2] It is a hypothetical portfolio since Norwegian mutual funds are not able to short-sell shares. Moreover, self-financing refers to zero investment trading strategy which is about longing past-winners and shorting past-losers. By adding spread portfolio, we are differentiating the estimates between top and bottom quintile.

[3] For the right tail of alphas bootstrapped p-value is equal to the percentage of simulated alphas that are greater than actual alphas and for the left tail of alphas the bootstrapped p-value is equal to the percentage of simulated alpha that is smaller than actual alpha.

portfolios is different from zero and whether it is generated by managerial skill (lack of skill) or luck to persistently over-perform (under-perform) over time**.**

The results in Table 6 indicate that there is no evidence of persistence in obtaining excess return over the benchmark for the top quintile of Norwegian mutual funds. The past-winner funds generate negative alpha in the holding periods up to 6 months for both 24-months and 36-months ranking period. This under-performance has a significant bootstrapped p-value at 5% level, suggesting that this negative return is due to the managers' lack of skill, not bad luck.

By looking at Panel A, we find a positive risk-adjusted return of 0.19% for the 24-month holding period; however, it is not statistically significant. The bootstrapped p-value implies that we fail to reject the null hypothesis that true excess return is equal to zero, indicating this excess return is produced by luck.

The results for bottom quintiles denote performance persistence for poor-performing funds in both Panels A and B. The excess return of the bottom quintile portfolio for all holding periods from 3 to 24 months is negative, with a significant bootstrapped p-value. It means that past-loser funds remain losers due to bad managerial skills. We reject the null hypothesis that persistence in underperforming is a result of bad luck. In panel B, for the shortest and the most extended holding period, 3 and 24 months, respectively, alpha is around -2%, which is the most significant magnitude among all holding periods.

Comparing the results in panel A and B, we find that for 24 months ranking period, the absolute value of negative alphas is greater than the value for the panel B for all holding periods in quintile 1 to 5. The significant excess returns in the spread quintile suggest that the difference between the best and worst funds' performance is meaningful; however, this trading strategy cannot be utilized by the Norwegian mutual funds' managers.

Table 6: Performance Persistence across Quintile Portfolios Formed on Lagged Alpha

This table shows annualized alphas and bootstrapped p-values from the Carhart 4-factor model for individual quintile portfolio (1 to 5), and for the spread portfolio (long in quintile 1 and short in quintile 5) using different lengths of ranking and holding periods. Quintile 1 represents the best-performing mutual funds portfolios, and quintile 5 represents the portfolios of the worst-performing mutual funds during the portfolio formation period. The portfolios are equally weighted of monthly returns, and the weights are rebalanced whenever a fund dies to eliminate survivorship bias. In Panel A (B), the alpha and bootstrap p-value of alpha are reported for portfolios of ranked mutual funds based on lagged 36 (24) months alpha and held for different holding periods (3, 6, 12, and 24-month periods). The bold p-values indicate significance at 5% confidence level. The procedure of ranking, portfolio forming, and holding repeated throughout the entire sample period (1998-2019).

	Holding Period							
	3 months		6 months		12 months		24 months	
Quintiles	Alpha	p-value	Alpha	p-value	Alpha	p-value	Alpha	p-value
Panel A: 36 months ranking period								
1	-1.10%	**0.00**	-1.19%	**0.00**	-0.36%	0.12	0.19%	0.27
2	-0.08%	0.40	-0.20%	0.22	-0.66%	**0.01**	-1.08%	**0.00**
3	-0.17%	0.24	0.09%	0.36	-0.25%	0.16	-0.60%	**0.01**
4	-0.01%	0.49	-0.42%	0.06	0.13%	0.32	-0.23%	0.23
5	-2.04%	**0.00**	-1.37%	**0.00**	-1.76%	**0.00**	-1.88%	**0.00**
1-5 spread	0.95%	**0.02**	0.19%	0.33	1.40%	**0.00**	2.08%	**0.00**
Panel B: 24 months ranking period								
1	-0.61%	**0.02**	-0.99%	**0.00**	-1.38%	**0.00**	-1.03%	**0.00**
2	-0.36%	**0.05**	-0.85%	**0.00**	-0.63%	**0.00**	-0.45%	**0.04**
3	-0.75%	**0.00**	0.01%	0.49	-0.43%	**0.03**	-0.21%	0.15
4	-0.53%	**0.04**	-1.32%	**0.00**	-0.46%	0.06	-0.66%	**0.02**
5	-2.32%	**0.00**	-1.50%	**0.00**	-1.91%	**0.00**	-2.30%	**0.00**
1-5 spread	1.71%	**0.00**	0.51%	0.09	0.53%	0.07	1.27%	**0.00**

Table 7 represents the results of the recursive portfolio method when funds are ranked and placed in quintiles based on the t-statistics of alphas. Aligned with the findings in Table 6, the negative risk-adjusted returns for the top quintile suggest that past-winners do not remain winners in the next ranking period, and consequently, there is no evidence of persistence of the best-performers in obtaining an excess return. The alpha of the top quintile portfolios for the 36-months ranking period and 24-months holding period is equal to 0.29%; however, it is not significant at 5% confidence level, meaning that we fail to reject that true excess return is equal to zero.

The outcome of bootstrapped simulation in Table 7 displays persistence among past losers both in Panel A and B, but the magnitude of the absolute value of negative alphas is greater with the 24 months ranking period. Besides, the alphas of the bottom quintile portfolio suggest that for the shortest and longest holding periods, 3 and 24 months, respectively, persistence in generating a negative risk-adjusted return is stronger. The p-value is significant at 5% confidence level, which implies that weak performance is not due to bad luck, but because of bad managerial performance.

Comparing Tables 6 and 7, we find the results for both ranking approaches are almost similar, suggesting that past-winner funds are not able to repeat their performance in the consequent period neither in the short nor in the long term. While past-losers, especially in the bottom quintile, exhibit persistence in generating negative risk-adjusted returns throughout all holding periods.

Table 7: Performance Persistence across Quintile Portfolios Formed on the t-statistics of Lagged Alpha

This table reports annualized alphas and bootstrapped p-values of the Carhart 4-factor model for each quintile portfolio (1 to 5), and for the spread portfolio (long in quintile 1 and short in quintile 5) using different lengths of ranking and holding periods. Quintile 1 represents the portfolios of best-performing mutual funds, and quintile 5 represents the portfolios of worst-performing mutual funds during the portfolio formation period. The portfolios are equally weighted of monthly returns, and the weights are readjusted whenever a fund disappears to avoid survivorship bias. In Panel A (B), the alpha and bootstrap p-value of alpha are reported for portfolios of ranked mutual funds based on t-statistic of lagged 36 (24) months alpha and held for different holding periods 3, 6, 12, and 24-month periods. The bold p-values indicate significance at 5% confidence level. The procedure of ranking, portfolio forming, and holding repeated throughout the entire sample period (1998-2019).

	Holding Period							
	3 months		6 months		12 months		24 months	
Quintiles	Alpha	p-value	Alpha	p-value	Alpha	p-value	Alpha	p-value
Panel A: 36 months ranking period								
1	-0.43%	0.07	-0.72%	**0.02**	-0.29%	0.18	0.29%	0.19
2	-0.15%	0.32	-0.09%	0.39	-0.10%	0.37	-0.40%	0.09
3	-0.33%	0.13	-0.61%	**0.02**	-0.55%	**0.04**	-0.64%	**0.01**
4	-1.32%	**0.00**	-0.80%	**0.01**	-1.21%	**0.00**	-1.23%	**0.00**
5	-1.13%	**0.00**	-0.90%	**0.01**	-0.81%	**0.02**	-1.54%	**0.00**
1-5 spread	0.70%	**0.03**	0.18%	0.29	0.52%	0.06	1.83%	**0.00**
Panel B: 24 months ranking period								
1	-0.27%	0.17	-0.82%	**0.00**	-1.00%	**0.00**	-1.38%	**0.00**
2	-1.12%	**0.00**	-0.78%	**0.00**	-0.55%	**0.02**	-0.63%	**0.00**
3	-0.58%	**0.02**	-0.12%	0.32	-0.60%	**0.01**	-0.43%	**0.03**
4	-0.23%	0.22	-1.75%	**0.00**	-1.14%	**0.00**	-0.46%	0.06
5	-2.23%	**0.00**	-1.03%	**0.00**	-1.49%	**0.00**	-1.91%	**0.00**
1-5 spread	1.96%	**0.00**	0.22%	0.25	0.49%	0.08	0.53%	0.07

5.5 Contingency Table

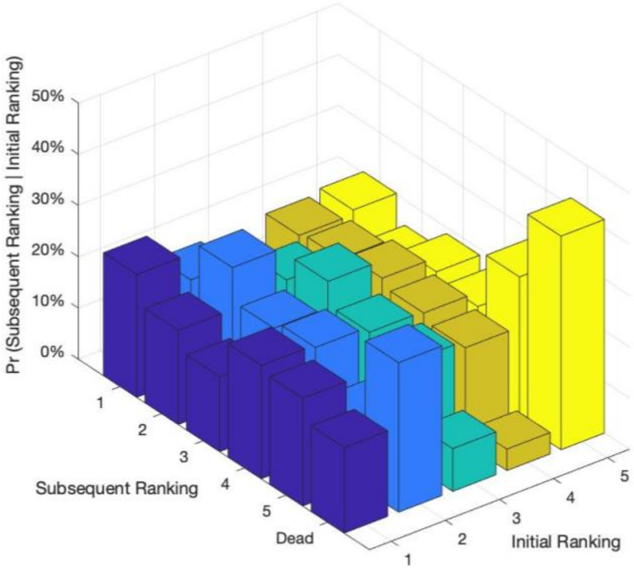
We employ a contingency table as a non-parametric approach to test if the track record of mutual funds returns can be used to predict performance in the subsequent period. Carhart (1997) suggests this approach to examine consistency in fund ranking by forming a contingency table of initial and following 12-month funds' rankings. We compare the net returns of 24 separate one-year periods and observe how funds move across the quintile portfolio from the initial period to the following one. The results of the constructed contingency table are illustrated in Figure 3, in which the bars for initial rank i and subsequent rank j denote the conditional probability of gaining a subsequent ranking of quintile j (or dying) given an initial ranking of quintile i.

Figure 3 indicates that the funds in the top, second, third, and bottom quintiles are more likely to repeat the performance in the following year; however, losers have the highest likelihood among all the funds to continue as losers in the subsequent period. Furthermore, last year losers have the highest probability of disappearing in the subsequent period. However, a high likelihood of funds disappearing in the second quintile might suggest that the perishing of the funds occurs randomly, and it happens due to unobserved factors rather than performance of a fund in the preceding period.

We also construct the contingency table in which mutual funds are ranked in quintiles based on the risk-adjusted returns to include the risk-taking level of funds' managers in assessing performance persistence. In this approach, the alphas of the Carhart 4-factor model are estimated for the subsequent 24-month and 36-month periods and then ranked in quintiles in each period. Similarly to the previous Figure 3, the bars indicate the Pr (rank j next interval | rank i last interval). The results for 24-month and 36-month intervals are presented in Figure 4A and 4B, respectively.

The findings for the risk-adjusted returns in the 24-month time intervals imply that loser funds in quintile four and five have a higher probability to either repeat the lousy performance or disappear. In contrast, winner funds have a higher likelihood of switching to the bottom quintile in the following period, which can be an indication of what Carhart described as gambling behavior by mutual funds (Carhart, 1997)

Figure 2: Contingency Table of Initial and Subsequent 12-month Performance Rankings



In this figure, the funds are ranked into quintile portfolios based on the lagged 12-month net returns from 1996 to 2019. These initial quintile rankings are compared with the subsequent quintile rankings. Funds that do not survive the entire subsequent year are categorized in a separate classification for dead funds. The bars in a cell (j, i) indicate conditional probability (rank j next period | rank i last period).

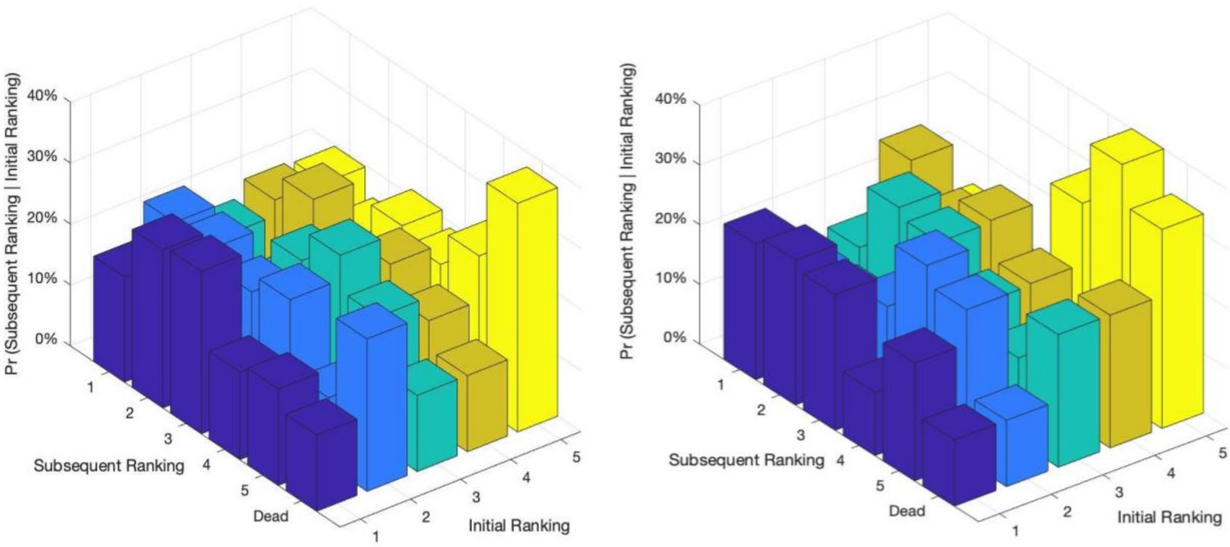
Similar to the 24-months risk-adjusted contingency table, the table for the 36-month alphas reveals performance persistence across worst-performing funds. It means that the worst-performers are more likely to either show poor performance or disappear in the following period. Contrastingly, top-ranked funds in the 36-months contingency table show less reverse performance than in the 24-months contingency table.

The findings of contingency tables of risk-adjusted returns are in line with the recursive portfolio approach that indicates past-losers remain losers, while winners are less likely to experience positive persistence over time. Compared to the analysis in the contingency table of risk-adjusted returns, the outcome of the raw net return contingency table shows that funds in quintiles 1 to 4 are almost equally likely to remain in the same quintile in the subsequent periods, and we can only observe a higher probability of performance persistence in the bottom quintile.

Figure 3: Contingency Table of Initial and Subsequent 24-month and 36-month Adjusted Returns Rankings

A: 24-month time interval

B: 36-month time interval



In this figure, the funds are ranked into quintiles based on alphas of the Carhart 4-factor model in the subsequent 24-month and 36-month time intervals. The initial quintile rankings are compared with the subsequent quintile rankings. Funds that do not survive the entire subsequent intervals are categorized in a separate classification for dead funds. The bars in the cell (j, i) denote the conditional probability of gaining a subsequent ranking of quintile j (or dying) given an initial ranking of quintile i.

5.6 Cross-product Ratio

While the contingency table illustrated negative performance persistence in the worst-performing quantiles of funds, cross-product ratio and the corresponding Z and Chi-squared statistics can not only detect persistence but also show whether the persistence is statistically significant.

In this section, we categorize a mutual fund every year as either a winner if the fund's raw annual return is above the median performer in the corresponding year or a loser if the raw annual return is below the median performer. After that, we calculate how many funds repeat or switch their performance in the following period and compile the number of WW, LL, WL, and LW throughout the whole sample period. By doing that, we obtain CPR, Z-statistics, and Chi-square statistics by following the methodology described previously. The results are presented in Table 8.

In most time periods and with the total sample period, we fail to reject the null hypothesis of CPR being equal to one, which means there is no indication of performance persistence.

However, in some periods, the null hypothesis is rejected, meaning there is performance persistence in several years. It is essential to point out that roughly half of the significant estimates are attributed to a reverse in performance. For instance, in years 1997-98 and 1998-99, most of the winner funds became losers, while most of the loser funds became winners; thus, the switch in performance is not a sign of performance persistence. Therefore, only in 4 out of 23 years, there is evidence of performance persistence. Nevertheless, since we are using raw returns, we cannot argue conclusively if the persistence comes from the managers' stock-picking abilities or lack thereof or simply because some managers take more systematic risks, which is rewarded with persistent returns.

Table 8: Performance Persistence Patterns on Raw Returns

The table reports cross-product ratios (CPR), Z-statistics (Z-statistic), and Chi-square statistics (χ^2) obtained by using annual raw returns net of fees. The total sample represents the values calculated for the whole time period from 1996 to 2019. The asterisks represent different significance levels – (***) is 1%, (**) is 5%, and (*) is 10%.

Time period	CPR	Z-statistic	χ^2
1996-97	2.70	1.29	1.71
1997-98	0.09	-2.92***	9.53***
1998-99	0.17	-2.57***	7.04***
1999-00	4.00	2.12**	4.67**
2000-01	6.48	2.85***	8.71***
2001-02	1.50	0.71	0.51
2002-03	0.34	-1.97**	3.96**
2003-04	0.54	-1.17	1.38
2004-05	2.38	1.65*	2.76
2005-06	1.30	0.51	0.26
2006-07	3.64	2.39**	5.93**
2007-08	1.00	0.00	0.00
2008-09	0.34	-2.04**	4.27**
2009-10	2.13	1.42	2.03
2010-11	0.25	-2.60***	7.01***
2011-12	0.52	-1.34	1.80
2012-13	1.94	1.34	1.80
2013-14	1.60	0.96	0.93
2014-15	6.86	3.40***	12.54***
2015-16	0.68	-0.75	0.57
2016-17	2.96	2.16**	4.77**
2017-18	0.59	-1.09	1.19
2018-19	2.25	1.72*	3.00*
Total sample	1.17	1.39	1.95

Since we would like to differentiate performance persistence better, we have also obtained CPR, Z-statistics, and Chi-square statistics using risk-adjusted returns, which is recommended in the literature (Brown & Goetzmann, 1995). We categorize a mutual fund every 24 and 36 months as either a winner if the fund's alpha is above the median alpha in the corresponding time period or a loser if the alpha is below the median estimate. The alphas are produced by applying the Carhart 4-factor model, which is the reason why we use at least 24 months in the analysis to produce meaningful estimates. After categorizing the funds, we follow the same methodology as described previously to obtain the statistics presented in Table 9.

When the returns are adjusted, most of the performance persistence disappears. In most time periods and with the total samples both for 24 and 36 months returns, we fail to reject the null hypothesis of CPR being equal to one; thus, there is no performance persistence. With the 24-months window, we reject the null for the period from 2000 to 2001, which indicates performance persistence for that period. Additionally, there is a reverse in performance from 2014 to 2015. With the 36-months window, we reject the null and claim performance persistence for a more extensive period from 1999 to 2001. The reversal performance also gets extended from 2014 to 2016.

Table 9: Performance Persistence Patterns with Alphas

The table reports cross-product ratios (CPR), Z-statistics (Z-statistic), and Chi-square statistics (χ^2) by using alphas produced by the Carhart 4-factor model with the time windows of 24 months (Panel A) and 36 months (Panel B). The total sample represents the values calculated for the whole time period from 1996 to 2019. The asterisks represent different significance levels – (***) is 1%, (**) is 5%, and (*) is 10%.

Panel A: 24 months

Time period	CPR	Z-statistic	χ^2
1996-97	0.29	-1.27	1.70
1998-99	1.00	0.00	0.00
2000-01	3.56	1.95**	3.94**
2002-03	1.38	0.57	0.32
2004-05	2.47	1.62	2.67
2006-07	0.94	-0.12	0.01
2008-09	1.85	1.16	1.36
2010-11	0.80	-0.41	0.17
2012-13	0.46	-1.38	1.93
2014-15	0.21	-2.66***	7.42***
2016-17	2.68	1.82*	3.38*
Total sample	1.10	0.55	0.31

Panel B: 36 months

Time period	CPR	Z-statistic	χ^2
1996-98	0.41	-1.04	1.11
1999-01	11.14	3.09***	11.07***
2002-04	1.97	1.15	1.34
2005-07	1.68	0.95	0.91
2008-10	2.23	1.46	2.15
2011-13	0.67	-0.70	0.49
2014-16	0.31	-1.96**	3.95**
Total sample	1.33	1.29	1.66

The analysis on risk-adjusted returns indicates one-time performance persistence in the whole sample period that lasts up to three years and another strong reversal in the performance that also lasts up to three years. Although there is persistence detected in one time period, the analysis of the total sample fails to display any significant performance persistence.

Compared to the analysis on raw returns, performance persistence with alphas seems to be less substantial, which can be partially explained by adjusting the returns for the risk factors and extending the time period.

6. Conclusion

We conduct a study on 99 actively managed Norwegian mutual funds that existed from 1996 to 2019 to examine if the mutual funds deliver in excess of the returns generated by passive benchmarks and whether the performance is persistent over time.

We find no excess returns when performance is evaluated against the Fama and French 3-factor and Carhart 4-factor models on the whole sample period. After aggregate performance assessment, we employ Fama and French technique (1997) to bootstrap the results we got and differentiate luck from skill in the returns.

We discover that when ranked by alpha estimates, the worst-performing funds in the sample have negative alphas at 5 or 10% significance level produced by the Carhart 4-factor model, meaning underperformance against the benchmark for those funds. The alpha estimates for the rest of the funds are indistinguishable from zero; therefore, there is no evidence of excess performance. When funds are sorted by t-statistics, which is generally perceived as a more accurate estimate, there is no abnormal performance among any funds. Therefore, the bootstrap analysis shows no evidence for managerial skill among Norwegian mutual funds.

To detect the presence of persistent performance, we follow Carhart's methodology (1997) to test yearly performance persistence on raw net returns. We find no indication of yearly persistence when performance is measured against the Fama and French and Carhart models. We also examine performance persistence using the recursive portfolio approach on risk-adjusted returns produced by the Carhart 4-factor model. Regardless of ranking methodology, the worst-performing quantiles of funds exhibit performance persistence up to 24 months. The results suggest that past-losers in the lowest-ranked quantiles persist in generating negative risk-adjusted returns both in the short and long term. In comparison, past winners fail to repeat their performance in the consequent periods. Since we also bootstrap the results, we find evidence that negative performance persistence is attributed to poor managerial skills rather than bad luck.

In addition to the tests mentioned above, we employ non-parametric techniques to test performance persistence. We construct contingency tables using raw net returns and risk-adjusted returns. The findings of the contingency tables are aligned with the previous results that losers remain losers in the subsequent time periods, while winners do not experience persistent positive performance over time. The results are similar for 12-, 24- and 36-month time intervals. Additionally, for all the contingency tables, the lowest-performing quintile of funds has the highest probability to perish in the subsequent period.

Another non-parametric method that we apply is a cross-product ratio. We obtain CPR with Z and Chi-squared statistics using raw net returns and risk-adjusted returns. With raw net returns, funds exhibit performance persistence in four different time periods, while with risk-adjusted returns and longer time interval, most of the persistence disappears. The only persistent returns are detected from 2000 to 2001 with the 24-month performance interval and from 1999 to 2001 with the 36-month performance interval. Nevertheless, we find no evidence of performance persistence when the tests are conducted on total samples for raw net and risk-adjusted returns.

All in all, we find no compelling evidence that the Norwegian fund managers generate positive excess returns for investors. The evidence shows poor managerial performance for the worst-performing funds. Regarding performance persistence, only the worst-performing funds continue to repeat their negative performance, which lasts up to two years. Many other funds, including the best-performing ones, reverse their positive performance in the subsequent period and deliver statistically significant negative performance when measured against the passive benchmark. The rest of the funds simply deliver zero values in excess returns.

Considering that we conduct our study on net returns, we cannot claim whether managers in Norway are able to produce excess returns before charging management fees. However, we can still say that investors do not receive the returns that are to be expected from actively managed mutual funds. Since the mutual funds' managers fail to deliver promised gains in excess of the passive benchmarks consistently, it is advisable to invest in a broad low-cost passive index rather than pay fees in vain.

Appendix 1: List of all mutual funds selected with the number of observations for each

Fund Name	Number of observations
Alfred Berg Aktiv	288
Alfred Berg Aktiv II	180
Alfred Berg Gambak	288
Alfred Berg Humanfond	240
Alfred Berg Norge +	195
Alfred Berg Norge Classic	288
Alfred Berg Norge Etisk	144
Alfred Berg Norge Inst	68
Arctic Norwegian Equities A	108
Arctic Norwegian Equities B	109
Arctic Norwegian Equities D	82
Arctic Norwegian Equities E	46
Arctic Norwegian Equities I	109
Arctic Norwegian Value Creation A	64
Arctic Norwegian Value Creation B	64
Arctic Norwegian Value Creation C	59
Arctic Norwegian Value Creation D	36
C WorldWide Norge	288
C WorldWide Norge III	212
Danske Invest Engros Norske A I R NOK I	54
Danske Invest Norge I	288
Danske Invest Norge II	288
Danske Invest Norge Vekst	288
Danske Invest Norske Aksjer Inst I	236
Danske Invest Norske Aksjer Inst II	157
Delphi Norge A	288
DNB Norge (Avanse I)	218
DNB Norge (Avanse II)	225
DNB Norge (I)	218
DNB Norge (III)	281
DNB Norge C	286
DNB Norge D	205
DNB Norge N	288
DNB Norge R	13
DNB Norge Selektiv	280
DNB Norge Selektiv (II)	212
DNB Norge Selektiv C	216
DNB Norge Selektiv E	288
DNB Norge Selektiv N	284
DNB Norge Selektiv R	13

Eika Egenkapitalbevis	216
Eika Norge	195
Eika SMB	185
FIRST Norge Fokus	13
Fondsfinans Norge	204
Formue Diversifiserte Norske Aksjer	110
FORTE Norge	105
FORTE Trønder	83
Globus Aktiv Acc	87
Globus Norge A/I	101
Globus Norge II Acc	93
Handelsbanken Norge	260
Handelsbanken Norge (A1 NOK)	288
Handelsbanken Norge (A10 NOK)	21
Holberg Norge A	228
KLP AksjeNorge	249
Landkreditt Norge	120
Landkreditt Utbytte A	82
Landkreditt Utbytte I	18
NB Aksjefond	205
Nordea 1 - Norwegian Equity AP NOK	50
Nordea 1 - Norwegian Equity BC NOK	74
Nordea 1 - Norwegian Equity BI NOK	24
Nordea 1 - Norwegian Equity BP NOK	265
Nordea 1 - Norwegian Equity E NOK	190
Nordea Avkastning	288
Nordea Kapital	288
Nordea Norge Pluss	104
Nordea SMB	212
Nordea Vekst	229
ODIN Norge A	288
ODIN Norge C	288
ODIN Norge D	288
ODIN Norge II	137
Pareto Aksje Norge A	207
Pareto Aksje Norge B	168
Pareto Aksje Norge C	53
Pareto Aksje Norge D	53
Pareto Aksje Norge I	219
Pareto Equity Edge D	41
Pareto Equity Edge P	62
PLUSS Aksje	276
PLUSS Markedsverdi	288

RF Aksjefond Acc	112
RF Plussfond Acc	52
Sbanken Framgang Sammen	47
SEB Norway Focus Fund C NOK	45
SEB Norway Focus Fund HNWC NOK	45
SEB Norway Focus Fund IC NOK	45
SEF FIRST SMB A NOK	116
Storebrand Aksje Innland	281
Storebrand Norge A	288
Storebrand Norge Fossilfri A	32
Storebrand Norge H	103
Storebrand Norge I	236
Storebrand Norge Institusjon	37
Storebrand Optima Norge B	219
Storebrand Verdi A	264
Storebrand Verdi N	21
Terra Norge	185

Appendix 2: Statistics of mutual funds used in the sample

Year	Total number of funds	Number of alive funds	Number of dead funds
1996	29	29	0
1997	34	34	0
1998	41	41	0
1999	42	42	0
2000	46	46	0
2001	49	49	0
2002	57	57	0
2003	59	59	0
2004	61	61	0
2005	62	62	0
2006	63	61	2
2007	61	60	1
2008	60	60	0
2009	60	59	1
2010	63	63	0
2011	67	67	0
2012	68	67	1
2013	70	68	2
2014	68	64	4
2015	66	64	2
2016	69	68	1
2017	70	69	1
2018	76	76	0
2019	76	72	4

References

Agarwal, V., & Naik, N. Y. (2000). Multi-period performance persistence analysis of hedge funds. The Journal of Financial and Quantitative Analysis, 35(3), 327–342. JSTOR. https://doi.org/10.2307/2676207

Berk, J. B., & Green, R. C. (2004). Mutual fund flows and performance in rational markets. Journal of Political Economy, 112(6), 1269–1295. https://doi.org/10.1086/424739

Bernt arne ødegaards financial data. (2020). Retrieved October 27, 2020, from http://finance.bi.no/~bernt/financial_data/index.html

Blake. D., Caulfield. T., Ioannidis. C., & Tonks. I. (2017). New evidence on mutual fund performance: A comparison of alternative bootstrap methods. Journal of Financial and Quantitative Analysis. 52(3). 1279–1299. https://doi.org/10.1017/S0022109017000229

Blake. D., & Timmermann. A. (1998). Mutual fund performance: Evidence from the uk. Review of Finance. 2(1). 57–77. https://doi.org/10.1023/A:1009729630606

Brown. S. J. Goetzmann. W., Ibbotson. R. G., & Ross. S. A. (1992a). Survivorship bias in performance studies. Review of Financial Studies. 5(4). 553–580. https://doi.org/10.1093/rfs/5.4.553

Brown. S. J.' Goetzmann. W., Ibbotson. R. G., & Ross. S. A. (1992b). Survivorship bias in performance studies. Review of Financial Studies. 5(4). 553–580. https://doi.org/10.1093/rfs/5.4.553

Brown. S. J., and Goetzmann. W. N. (1995). Performance persistence. The Journal of Finance. 50(2). 679–698. JSTOR. https://doi.org/10.2307/2329424

Carhart. M. M. (1997). On persistence in mutual fund performance. The Journal of Finance. 52(1). 57–82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Carpenter. J. N., & Lynch. A. W. (1999). Survivorship bias and attrition effects in measures of performance persistence. Journal of Financial Economics. 54(3). 337–374. https://doi.org/10.1016/S0304-405X(99)00040-9

Christensen. R. (1990). Log-linear models. Springer New York. https://doi.org/10.1007/978-1-4757-4111-7

Elton, E. J. (1993). Efficiency with costly information: A reinterpretation of evidence from managed portfolios. Review of Financial Studies, 6(1), 1–22.

Elton. E. J., Gruber. M. J., & Blake. C. R. (1995). Fundamental economic variables. expected returns. and bond fund performance. The Journal of Finance. 50(4). 1229–1256. https://doi.org/10.1111/j.1540-6261.1995.tb04056.x

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. https://doi.org/10.1016/0304-405X(93)90023-5

Fama. E. F., & French. K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. The Journal of Finance. 65(5). 1915–1947. https://doi.org/10.1111/j.1540-6261.2010.01598.x

Grinblatt. M., & Titman. S. (1992). The persistence of mutual fund performance. The Journal of Finance. 47(5). 1977–1984. https://doi.org/10.1111/j.1540-6261.1992.tb04692.x

Hendricks. D., Patel. J., & Zeckhauser. R. (1993). Hot hands in mutual funds: Short-run persistence of relative performance. 1974-1988. The Journal of Finance. 48(1). 93–130. https://doi.org/10.1111/j.1540-6261.1993.tb04703.x

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. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jensen. M. C. (1968). The performance of mutual funds in the period 1945-1964. The Journal of Finance. 23(2). 389–416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x

Kosowski. R., Timmermann. A., Wermers. R., & White. H. (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. The Journal of Finance. 61(6). 2551–2595. https://doi.org/10.1111/j.1540-6261.2006.01015.x

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. JSTOR. https://doi.org/10.2307/2328251

Malkiel. B. G. (1995). Returns from investing in equity mutual funds 1971 to 1991. The Journal of Finance. 50(2). 549–572. https://doi.org/10.1111/j.1540-6261.1995.tb04795.x

Morningstar investment glossary. (n.d.). Retrieved July 2. 2020. from http://www.morningstar.com/InvGlossary/

Næs. R.. Skjeltorp. J.. & Ødegaard. B. A. (2009). What factors affect the Oslo stock exchange? (2009/33; UiS Working Papers in Economics and Finance). University of Stavanger.

Newey. W., & West. K. (1987). A Simple. Positive Semi-Definite. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica. 55(3). 703. doi: 10.2307/1913610

Nibor – the Norwegian Interbank Offered Rate. (2020). Retrieved from https://www.finansnorge.no/en/interest-rates/nibor---the-norwegian-interbank-offered-rate/

Oslo Børs – Equity Indices – Index Methodology. (2020). Oslo Børs.

Sørensen, L. Q. (2009). Mutual fund performance at the Oslo stock exchange. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1488745