Norwegian School of Economics

Bergen, Fall, 2016

Investing in sustainability

The risk-adjusted performance of European mutual funds committed to sustainable and responsible investing

Adrian Krigsvoll Auran and Alexander Kristiansen

Supervisor: Nils Friewald

Mastertheis in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This paper examines the relationship between sustainability and traditional financial aspects. Sustainable development has manifested itself to financial markets and the newly launched Morningstar Sustainability Rating serves investors with quantifiable and objective measures of mutual funds sustainability. We use this measure to infer causality between financial performance and investment style and find no statistical evidence that there exist a risk-adjusted performance advantage or disadvantage from investing in sustainability. The findings imply that there is no additional cost related to investing in sustainable mutual funds, which might be interesting for value-driven investors. The funds categorized as the most sustainable are found to be more sensitive to market and large capitalization stock returns relative to the funds categorized as being the least sustainable. Our findings are robust for a range of sustainability definitions, management fees and transaction cost.

Contents

1. Introduction

Social and environmental awareness is imperative to ensure a sustainable global economy. According to Intergovernmental Panel on Climate Change (IPCC, 2014) a key transitory factor into sustainable development is investment opportunities to secure future returns in which the financial system facilitates future economic well-being. Sustainable and responsible investing (SRI) involves including non-financial considerations, such as environmental, social, and governance (ESG) factors, to a traditional investment process. Consequently, sustainable and responsible investors constrain themselves by investing in a subset of the investable universe using various investment screens. Drawing on Modern Portfolio Theory (Markowitz, 1952), imposing constraints on the portfolio construction process will likely come at the cost of less efficient portfolios and thus inferior performance. Advocates of SRI however, argue that the social and environmental benefits of integrating ESG considerations into the investment process more than offsets the loss of portfolio efficiency.

SRI is experiencing rapid growth reflecting the increasing investor awareness to sustainability issues. Academic interest has followed the growth and there is now an extensive literature investigating the relationship between sustainable and financial performance. Empirical findings are somewhat conflicting, although the majority of studies find no evidence of significant differences between sustainable funds and conventional peers (e.g. Bauer et al., 2005). Despite several studies undertaking the financial aspect of SRI, the appropriate tools for assessing sustainability so far have been lacking. With the recent introduction of standardized and objective measures of sustainability from independent data providers, SRI analysis are likely to increase in terms of quality.

The objective of this study is to examine whether European open-end mutual funds categorized as the most sustainable differentiate themselves from the least sustainable in terms of risk-adjusted performance and investment style. Fund level SRI-analysis mitigate the problem arising from determining the direction of causality typical for firm level analysis. This is related to whether good financial performance leads to good responsible practices or the other way around. Johnsen and Gjølberg (2003) state that although firm-level studies have its value, more solid conclusions regarding fund performance demand portfolio-level analysis. We use the newly launched Morningstar Sustainability Rating to

categorize funds and construct two equally weighted portfolios of returns - one of high-sustainability funds and one of low-sustainability funds. Using CAPM (Lintner, 1962; Sharpe, 1964; Mossin, 1966; Black, 1972), Fama-French three-factor model (Fama and French, 1993) and Carhart four-factor model (1997), we control for common risk factors in order to infer differences in risk-adjusted abnormal returns and factor loadings. To the best of our knowledge, this study is one of the first to apply Morningstar Sustainability Rating to evaluate the performance of sustainable mutual funds. Our dataset renders more objective classifications, enabling us to move forward from typical SRI-performance analysis based on funds' self-reported sustainability assessments and matched conventional benchmarks.

Our results are easy to summarize. High-sustainability funds yield no significant difference in returns relative to low-sustainability funds after controlling for market, size, value and momentum risk factors. In terms of investment style differences, we find that high-sustainability funds are more exposed to market risk and invest more in large capitalization stocks relative to low-sustainability funds. This implies that high-sustainability funds invest relatively more in cyclical stocks that are more sensitive to market returns. They are also less exposed to small cap stocks which are generally considered to be riskier than large cap stocks. To highlight the advantage of using non-dichotomous measures of sustainability in SRI-analysis we ensure robustness by characterizing our results even further by also using different sustainability definitions.

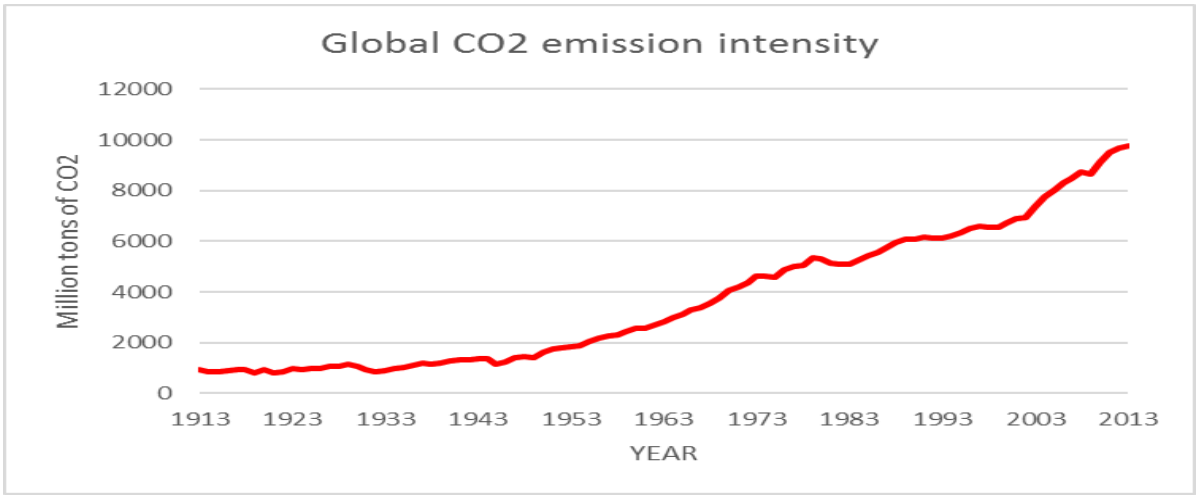
2. Research rationale

2.1 Global sustainability problems

Our planet is currently facing enormous sustainability problems. Business and commerce has played a significant role in the problematic trends following overconsumption as a consequence of the exponential population growth. It is where consumption demand meets production supply business and commerce has the potential to help solve these problems. The global population has increased by more than 350 percent in the 20th century alone as we now have surpassed 7 billion inhabitants. Even though the extreme population growth in the last 100 years is expected to decline, the population is projected to reach 9 billion by 2045 which in turn will characterize the consequences of overpopulation even more (Un.org, 2016). Overpopulation is moving us towards a dreary future due to the consequences of *overconsumption* and *environmental degradation* (Ehrlich and Ehrlich, 1990). Overconsumption here refers to the gradual depletion of limited resources, while environmental degradation refers to how environmental capacities degrade due to climate change and global warming. The paradox is that the population growth is causing consume of so many scarce resources that it is impossible to sustain future generations. At the same time the production of these goods are causing direct harm to both *social* and *environmental sustainability*.

2.1.1 **Environmental sustainability**

Figure 2.1: Development in CO2 emission intensity last 100 years



Source: Boden, T.A., G. Marland, and R.J. Andres. 2016. Global, Regional, and National Fossil-Fuel CO2 Emissions.

Climate change is causing global temperatures to rise which can have catastrophic implications including extreme weather, ice melting in the artic and rising sea levels. Scientific consensus on global warming is that it is human caused (Cook et al., 2016). Some argue a causal relationship between population growth and increasing change in climate (Houghton, 2004). The greatest driver of observed climate change is greenhouse gas emissions associated with human actives in which 65 percent of the total global emission stem from CO2 emissions related to combustion of fossil fuel and industrial processes (IPCC, 2014). The increasing emission intensity of carbon dioxide shown in figure 2.1 represents the environmental harm that is happening as a consequence of human consumption. Production of electricity and heat is the number one source of greenhouse gas emissions (25 percent of total greenhouse gas emission) and illustrates how the paradoxal production of consumer goods like electricity is by causing serious harm to the environment by burning fossil fuel, while at the same time depleting fossil fuel as a finite resource (IPCC, 2014).

2.1.2 **Social sustainability**

Social sustainability relates to how corporations identify and manage negative and positive externalities that directly or indirectly affect employees and stakeholders (UNglobalcompact.org, 2016). Business' have a responsibility in supporting and obliging to international human rights standards to ensure proper working conditions. Even still, child labor and hazardous workplaces in large, influential multinational corporations are being uncovered and brought to the media's attention ever so often. When talking about a sustainable future we are interested in determinants that facilitates the social and environmental pillars of sustainable development. Sustainable development involves maintaining humanities welfare without exhausting Earth's resources or damaging its natural system.

3. Sustainable and responsible investing

We define sustainable and responsible investing (SRI) as an investment approach that considers environmental, social and governance factors throughout the investment process. This definition is consistent with that of the European Social Investment Forum (Eurosif, 2016). For the purpose of this paper, we will use an inclusive definition of SRI without drawing distinctions between this and related terms such as responsible investing, socially responsible investing or ethical investing. These are collectively referred to as sustainable and responsible investing or SRI.

Eurosif (2016) classifies SRI-strategies in seven distinct categories: exclusions, norms-based screening, best-in-class selection, sustainability themed investments, ESG integration, engagement and voting, and impact investing. We group these investment approaches into three main bodies, namely negative screening, positive screening and engagement.

3.1.1 **Negative screening**

Negative screening or exclusions is a strategy that involves eliminating certain companies or sectors from the investment universe based on specific ESG criteria. This approach remains dominant and covers more assets than any other SRI strategy with respect to assets under management. Another type of negative screening is the norms-based approach, which involves screening of investments against minimum standards of business practice based on international norms or principles. These norms are typically based on conventions set by organizations or institutions such as OECD or UN (Eurosif, 2016).

3.1.2 **Positive screening**

Positive screening is an umbrella term that transcends several SRI-strategies. Instead of excluding companies engaged in controversial activities, a positive screening approach includes sectors, companies or projects selected for positive ESG performance. Best-in-class selection involves identifying the leading or best-performing companies within a sector using ESG criteria, and including these in the investable universe. ESG integration is the explicit inclusion by investment managers of environmental, social and governance factors into traditional financial analysis and investment decisions. This strategy is gaining momentum, not only in Europe but also across the globe. This is based on the conviction that

it is imperative to reconsider traditional financial analysis and look at companies from a holistic point of view (Eurosif 2016). Sustainability themed investing are typically investments in themes or assets specifically related to sustainability, for example clean energy, green technology or sustainable agriculture. Lastly, impact investing intentionally seeks to create both financial return and positive social or environmental impact, and is another distinct way to channel funding to social organizations or companies that seek to handle specific social challenges through market mechanisms.

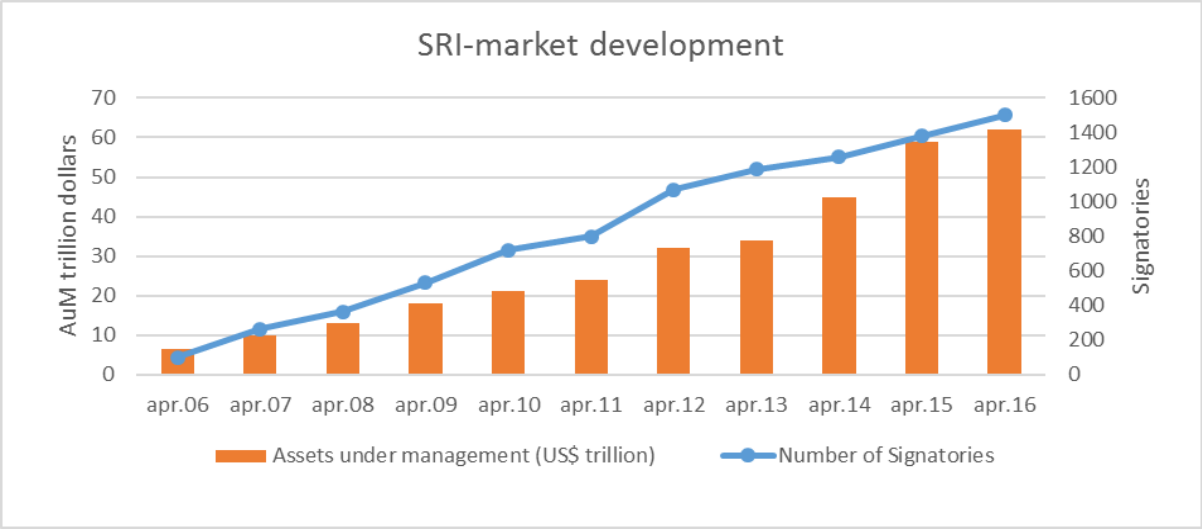
3.1.3 **Engagement**

Engagement involves taking on the role of active ownership, employing shareholder power through voting of shares and direct corporate engagement on ESG matters. It has a very strong link to fiduciary duty, as it is driven in large part by the view that shareholders are steward of assets who are accountable to their beneficiaries for how they manage those assets (Eurosif 2016).

3.2 The market of SRI

SRI is a fast growing industry especially in European and US capital markets. At the outset of 2016, almost 9 trillion dollars were managed using SRI strategies in the US. This represents more than one out of five dollars of assets under professional management, and is an increase of more than 30 percent from beginning of 2014 (The US SIF Foundation, 2016). In Europe, just over 11 trillion euro were managed using SRI-strategies in 2015, which is an increase of 12 percent from 2013 (Eurosif, 2016). The increased popularity of SRI is driven by increased awareness of institutional and individual investors to ESG issues. As a result, the number of mutual funds and exchange-traded funds catering to these investors has mushroomed in recent years. The United Nations-supported Principles of Responsible Investment was initially launched in 2006 with 100 *signatories* representing 6.5 trillion dollars of AUM. Signatories represent large institutional investors and investment managers who commit to six principles for responsible investment. Among them are 11 of the 15 largest investments managers in the world including BlackRock, Vanguard, JPMorgan, Goldman Sachs and PIMCO (Unpri.org, 2016).

Figure 3.1: UNPRI's contribution to development of SRI-market



Source: Unpri.org, 2016

Interest in SRI appears to be growing, particularly among women and millennials – two groups who are rapidly becoming more influential investment decision-makers. A 2015 survey conducted by the Morgan Stanley Institute for Sustainable Investing find that female investors are nearly twice as likely as male investors to consider ESG factors when making investment decisions. Moreover, millennial investors are twice as likely to make sustainable investment decisions as other investors. (Morgan Stanley Institute for Sustainable Investing, 2015). With these demographic trends providing a tailwind, SRI is likely to continue moving into the mainstream in the coming years.

4. Theoretical background

4.1 Firm-level analysis

Johnsen and Gjølberg (2003) highlight the presence and dangers of "excellent cases" in firm-level studies, i.e. studies often tend to focus on companies that appear to be or are sustainable and historically also has performed very well. The companies that are otherwise similar but not as good performers, are to a greater extent neglected. Reviewing firm-level research gives support to the hypothesis that there is a heavy overrepresentation of excellent companies. By using a wide range of key words related to sustainability we search the most reputable financial journals and find a lot of studies that support the hypothesis that sustainability contributes to increased corporate performance (Cremers and Ferrell, 2014; Core et al., 2006; Deng et al., 2013; Edmans, 2011; Hong and Kacperczyk, 2009). We find fewer studies (Di Giuli and Kostovetsky, 2014) that supports to disprove this hypothesis. Fund-level SRI-analysis mitigate the problem arising from determining the direction of causality typical for firm-level analysis. Does good measures related to financial performance cause good measures of responsibility or the other way around. Johnsen and Gjølberg (2003) state that although firm-level studies have its value, more solid conclusions regarding fund performance demand portfolio-level analysis.

4.2 The financial system

According to IPCC (2014), a key transitory factor into sustainable development is the role of the financial system. The financial sector represents a well-functioning economy in which capital is allocated and risk is being managed in an efficient way. By enabling risk insurance and investment opportunities future returns are secured through the financial system, which is laying the foundation for future well-being. This is in accordance with the definition of sustainable development: current needs being met without compromising the needs of the future (United Nations, 1987). The increasing demand for sustainable development has therefore manifested itself in the financial system with a fast rising market of sustainable and responsible investing (SRI) where Europe are global leaders.

4.2.1 Possible screening implications for mutual funds

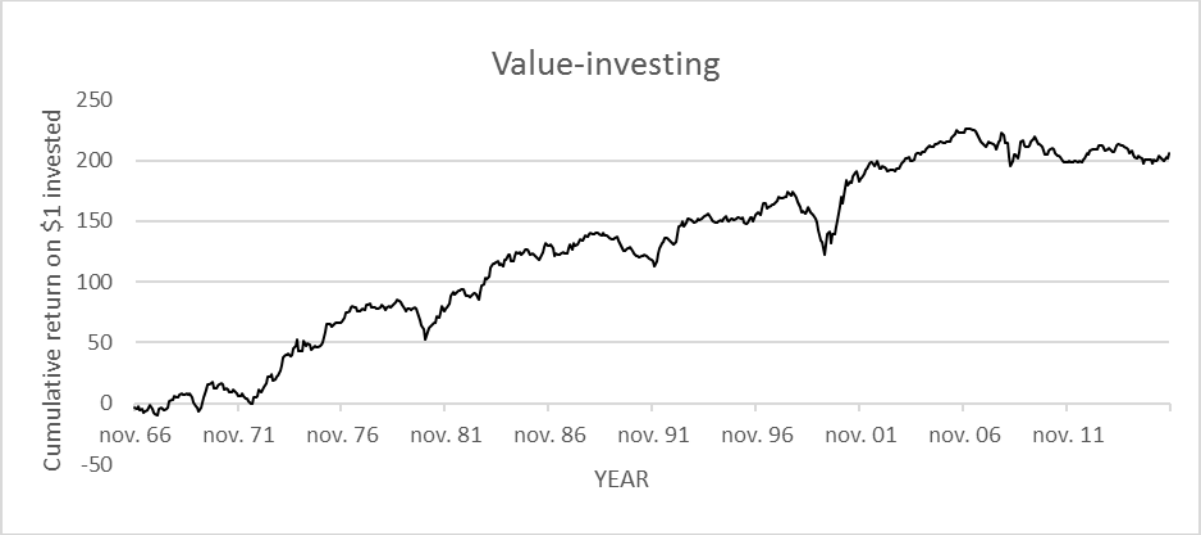
Modern Portfolio Theory (Markowitz, 1952) is a Nobel award winning mathematical framework for constructing a portfolio of assets where the expected return is maximized given a certain level of risk. This relationship indicates a mean-variance optimal portfolio. Imposing constraints on the optimization process introduce a possibility of omitting efficient assets from the investment universe. According to Ang (2015) the optimal mean-variance portfolio is never constrained. If the investor is lucky, the most rewarding risk-reward relationship will stay unaffected by adding constraints. If ESG factors are negatively related to financial performance, then screening for these factors may cause the portfolio to underperform. Hong and Kacperczyk (2009) find that "sin" stocks, i.e. companies involved in in the production of tobacco, gambling and alcohol, consistently outperform comparable stocks.

Investors interpretation of efficient markets should be determinate of whether or not they should invest in active management. If the investor believes the efficient market hypothesis, or EMH (Fama, 1970), to be true, it is not possible to beat the market. If investors believe there is reason not to trust the EMH, then it is possible to beat the market by investing in active management. If market participants systematically underestimate the benefits or overestimate the costs of SRI, then the expected return of sustainable companies might be consistently higher (Statman and Glushkov, 2009). Edmans (2011) find that the market undervalues intangibles, and firms with greater employment practices consistently deliver superior earnings performance. Pedersen (2015) view markets as neither perfectly efficient nor completely inefficient. By calling the market *efficiently inefficient* he suggests that markets are inefficient enough so that active asset managers are compensated for their management cost and efficient enough so that additional active investing is discouraged. The management fee represents income for asset managers and cost for investors and the size of the fee should therefore be seen in relation to the amount of effective management provided as well as management quality (Pedersen, 2015).

Value stocks have historically outperformed growth stocks. Value stocks are characterized by a low market price relative to book value of equity, while growth stocks are expensive relative to their book value. Stattman (1980) was the first to find book values to be positively correlated with stock returns. The value-strategy (Fama and French, 1993) involves going

long in value stocks (high book-to-market) and short in growth stocks (low book-to-market). Figure 4.1 illustrates the performance advantage of value-investing by plotting the cumulative returns of a one-dollar investment over the last 50 years.

Figure 4.1: Cumulative return 50-year value-strategy



Source: Fama-French HML Factor (Mba.tuck.dartmouth.edu, 2016)

Negative screening strategies typically avoid growth stocks while positive screening strategies typically favor value stocks. Top exclusion criteria include growth stocks like tobacco and alcohol (Eurosif, 2013). European industry book-to-market ratios for aggregated tobacco and alcohol stocks are 0.1 and 0.3, respectively (People.stern.nyu.edu, 2016). Alternatively, most popular industries of positive screening strategies include value stocks like waste services and agriculture (Eurosif, 2016) with industry book-to-market ratios of 1.7 and 2, respectively (People.stern.nyu.edu, 2016).

5. Literature review

5.1 Performance

5.1.1 Negative perspective

Renneboog et al. (2008) study the impact of ethics and stakeholder governance on fund performance in a global context. Their hypothesis state that investors in sustainable funds pay a price that makes them underperform compared to conventional funds and benchmarks. To categorize SRI-funds the authors rely on self-reporting by creating a list of funds that are labelled with certain words like ethical, environmental, sustainable and similar tags in Standard & Poor's Fund Service. Findings suggest that sustainable funds in many European, North-American and Asia-Pacific countries strongly underperform against domestic benchmark portfolios. Although alpha is lower for all sustainable funds compared to conventional funds, alpha is only statistically significant for a small a group of countries, namely France, Ireland, Sweden, Japan and Singapore. Sustainable funds in the US and the UK show no statistically significant difference in alpha compared to conventional funds and benchmark portfolios.

Johnsen and Gjølberg (2007) is an update of their 2003 study by request from the Ministry of Finance. Their conclusions suggest a significant degree of variability within the three performance perspectives but results strongly points in the direction of having a negative impact on risk-adjusted performance for SRI-funds through the upward trending economic cycle from 2003 to 2007. They attribute this to SRI-funds and indices showing higher idiosyncratic risk than conventional funds and indices.

5.1.2 Neutral perspective

Bauer et al. (2005) look at the differences in risk-adjusted returns and investment style between a total of 103 ethical labelled mutual funds and a matched sample of 4 384 conventional funds. The ethical fund sample is created using lists of funds that according to independent providers of financial information like EIRIS (UK) use ethical screens in their mandate. This approach work in our opinion better than the self-reporting system, but the authors does not provide any content or boundaries to what is defined as ethical. By

employing the Carhart four-factor model they find no statistically significant differences in risk-adjusted returns.

Statman (2000) takes several approaches to investigate how ethical investing perform compared to conventional ways of investing. The author aims to map how the Domini Social Index performs compared to the S&P 500 index. Statman also looks at how individual SRI-funds perform against mentioned indices in addition to matched conventional funds. The SRI-fund sample is created using a list over socially conscious mutual funds that according to Morningstar are funds that impose major socially responsible constraints on investment activity. It is worth noting that this list is not the same as Morningstar Sustainability Rating. Alpha works as performance measure in the one factor model approach. Comparing funds, Statman finds that SRI-funds on average perform better than conventional funds. The DSI performed similar to the S&P 500, giving slightly lower risk adjusted returns although not statistically significant. Lastly, the SRI-funds did on average perform worse than both the S&P 500 and DSI. Interestingly, the SRI-funds tracked the S&P 500 better than DSI.

5.1.3 **Positive perspective**

According to Johnsen and Gjølberg (2003) there might be a non-symmetrical relationship between risk-adjusted returns and the economic cycle. Although their general conclusion strongly points toward a neutral perspective they find that sustainable funds seem to do better when the economic cycle is upward trending or normal/neutral and that they lose in economic downturns. The authors states that this phenomenon is not due to the ethical constraints on the portfolio, but rather that ethical funds are non-normal in pricing downside risk and this is revealed only in abnormal situations. Opposite to this conclusion, Nofsinger and Varma (2014) find that socially responsible mutual funds outperform conventional funds during downturns in the economic cycle. Thus, findings imply a smaller downside risk compared to conventional funds. This asymmetry is attractive for investors with a Prospect Theory (Kahneman and Tversky, 1979) utility function because a given loss is felt more strongly than an equivalent sized gain. The authors thus reject the hypothesis that SRI-funds are attractive for certain investors because of the impact of externalities occurring from ethical portfolio constraints on the investors utility function. This view is quite the contrary from a lot of research in the area which in turn makes it one of the more interesting studies in favour of the positive perspective.

In Nofsinger and Varma (2014) the SRI-fund sample is constructed using several sources. Similar to Statman (2000) the authors use Morningstar's list over SRI-funds, in addition to looking at funds self-reporting. They obtain a relatively large sample size of 240 SRI-funds for the sample period of 2000 - 2011. Again similar to Statman (2000), the authors pair the SRI-funds with three conventional funds per observation. Concluding results are obtained looking at alpha for the three common factor models; CAPM, Fama-French 3 factor and Carhart 4 factor. Two major economic downturns are identified. The dotcom bubble (2000-2002) being the first and the global financial crisis (2007-2009) being the second. For the full sample period results are consistent with other research. The historic average return on SRI-funds are smaller than conventional peers. Alphas are not statistically significant, negative and smaller for all models except CAPM. During times of major financial distress average return on SRI-funds are better than conventional funds. This represents according to the authors the pricing of reduction in downside risk. The benefit is gained during distress and is, on average an increase of 96 basis point compared to the cost of 79 basis points. In a non-crisis scenario alpha for all models are significantly smaller for SRI-funds. In crisis all alphas are positive and larger for SRI-funds, although not of statistically significant value.

5.2 Investment style

5.2.1 **Market loading**

Renneboog et al. (2008) finds a small negative statistically significant difference in market loading with the Carhart four-factor model. This indicates that the SRI-funds are slightly less sensitive to the market returns than conventional funds. Although, there are some countries that exhibit a large variability from the average: Luxemburg, Sweden, Canada and Singapore. In an international perspective Bauer et al. (2005) find a small, negative and statistically significant difference on market beta using both CAPM and Carhart four-factor model. For SRI-funds this implies an overall less sensitivity to market returns. Nofsinger and Varma (2014) finds a small, positive and statistically significant difference implying that SRI-funds tend to be a little more sensitive to market returns than conventional funds.

5.2.2 **Size loading**

Comparing SRI-funds with conventional funds, Renneboog et al. (2008) find that Germany and the UK have larger exposure to small cap stocks. While the US, Canada and Japan invest

considerably more in large caps. For Bauer et al. (2005) the results are more consistent. Germany, the UK and Internationally load a positively, statistically significant difference from conventional funds. Only the US have a negative difference in loading compared to conventional funds. This implies that German and UK SRI-funds are highly exposed to small cap stocks while the US is more involved in large cap stocks in their portfolios. In the Nofsinger and Varma study (2014) they find a non-significant loading zero for SRI-funds and a significant beta for conventional funds.

5.2.3 **Value loading**

Renneboog et al. (2008) find that SRI-funds in Norway, Canada and Japan have a larger exposure to value stocks than conventional funds. Generally, most countries in the study exhibit a negative loading which in turn implies a tilting towards growth stocks, although not statistically significant Bauer et al. (2005) find that all ethical fund portfolios are more growth-oriented than conventional funds. Germany, the UK, International and the US load significantly different from conventional funds. Nofsinger and Varma (2014) conclude that SRI-funds load a little more on the HML factor than conventional funds.

5.2.4 **Momentum loading**

According to Renneboog, Ter Horst and Zhang (2008) SRI-funds load, on average less on the momentum factor than conventional funds. The exceptions are Norway, Netherlands, Switzerland and Australia, whom all have a non-significant positive difference compared to conventional funds. Their findings are supported by Nofsinger and Varma (2014). They find that SRI-funds load with a negative difference compared to conventional funds. In Bauer et al. (2005) the results are inconclusive. Internationally, SRI-funds load higher on momentum for SRI and conventional funds. Germany, the UK and the US load less.

6. Data

Morningstar is a leading provider of independent investment research and provide extensive data on investment vehicles such as mutual funds. Mid-2016 Morningstar launched its Morningstar Sustainability Rating (MSR) providing financial speculators with ground-breaking objective measures of sustainability performance. Directed at mutual funds, Morningstar provide absolute and relative ranking of how well a fund manage the three ESG dimensions of sustainability. This implies a quantification of sustainability performance thus serving as a measurable standard of the ESG dimensions. This does not only benefit investors looking to invest in sustainability, but also researchers that want to investigate if sustainability investing is profitable.

6.1 Categorizing sustainable funds

The Morningstar Sustainability Rating is derived from the Morningstar Portfolio Sustainability Score calculated as following (Morningstar.no, 2016):

Portfolio Sustainability Score = Portfolio ESG Score – Portfolio Controversy Deduction

Portfolio ESG Score is the asset-weighted average of normalized company-level ESG scores and is a measure of how well a company is addressing ESG issues based on a series of indicators related to preparedness, disclosure and performance. Portfolio Controversy Deduction refers to companies ESG-related incidents and is assessed by the impact on the environment, society and the risk for the companies itself. The two measures combine to display a score between 0 and 100. A high score indicates that a fund has its majority of assets invested in stocks that has a high ESG score according to the methodology of Sustainalytics - a leading global independent data provider specializing in sustainability research.

6.1.1 **Sustainalytics**

Sustainalytics' ratings are measures of how well companies proactively manage environmental, social and governance (ESG) issues that are considered material. Materiality refers to the importunateness certain sustainability issues have for the individual companies. For each of the individual ESG-factors, Sustainalytics apply a three dimensional assessment on (Sustainalytics, 2016): a) *preparedness*, b) *disclosure* and c) *performance*. Preparedness refers to how companies' management systems, programs and policies are designed to

manage ESG risk. The disclosure dimension refers to the level of transparency in companies' internal and external reporting routines and how the company comply with best international standards. Performance is measured with two separate indicators. Qualitative indicators are assessments of companies' level of involvement in controversial incidents and is rated on a scale based on the impact on society and the environment. Quantitative indicators on the other hand are concerned with easily measurable metrics like for example carbon release. Covering all three dimensions, Sustainalytics apply more than 70 core and industry specific indicators that are analysed, scored and weighed to determine a company-level ESG score. Sustainalytics provide insight into their four step research process; 1) Obtaining data on company via media, own disclosure and NGO's, 2) analysis of data within the 3x3 dimensional framework, 3) peer review and 4) company contact to receive feedback and update data.

6.1.2 **Defining other fund criteria**

We wish to investigate open-end mutual funds with equity exposure of minimum 80 percent. According to The Norwegian Fund and Asset Management Association, the requirement for a mutual fund to be classified as an equity fund is that it normally has an 80-100 percent exposure to equity (Vff.no, 2016). To increase the comparative value of our research to previous research we define the investment universe to the geographical area of Europe. Because of the sheer size of UK financial markets Morningstar Direct separate it from the rest of Europe and we thus say that we require a minimum of 50 percent of portfolio invested in either European developed markets (excluding emerging markets defined as Hungary, Poland, Romania and Russia) or 50 percent in the UK. We define the fund category and limit the ultimate sample to funds that 1) has a minimum of 80 percent holdings in equity, 2) a minimum of 50 percent of investments done in either Europe or the UK, 3) is marketed to Norwegian investors and 4) has a Morningstar Sustainability Rating.

6.1.3 **Morningstar Sustainability Rating**

Funds are given the Morningstar Sustainability Rating within the defined category an absolute score, a relative percentage ranking and a descriptive rating. The rating is based on the Portfolio Sustainability Score given and is normally distributed within the category as defined in the section Fund Category funds.

Table 6.1: Normally distributed sustainability ratings in fund category

Percentage Rank	Descriptive score	Category
Top 10 %	High	128 (7%)
Next 22.5 %	Above Average	343 (20%)
Next 35 %	Average	522 (30%)
Next 22.5 %	Below Average	508 (29%)
Bottom 10 %	Low	239 (14%)

We wish to investigate how high-sustainability funds perform against low-sustainability funds and define our *high* and *low* portfolios according to Morningstar's grouping of the top 10 and bottom 10 percent ranking funds, respectively. We have 128 high-sustainability and 239 low-sustainability funds that we are going to examine further.

6.2 Sample period

6.2.1 **Persistence of sustainability rating**

MSR is a cross-sectional variable representing how sustainable a fund is per today, i.e., it does not have the same time series properties as past returns. Studies using such longitudinal data do not exist as of today, mainly because such data sets do not exist. It is worth mentioning that this is a limitation that all previous research is subject to and our research does not fall short of others because of it. The reason we address this is because we need to evaluate to which point in the past the cross-sectional nature of high and low categorization might be valid for. Wimmer (2012) examines the persistence of ESG-scores in mutual funds and find statistical evidence that a funds ESG-score is persistent over a two-year period and that high and low scoring funds in terms of ESG tend to have scores converging towards the mean within three years. It is worth noting that the author is using much wider percentile distributions to categorize funds as high or low with top and bottom 25 percentile, respectively. Our top and bottom 10 percentile categorizations for high and low work as more robust distributions and we find it more likely that these categories have and will implement ESG-investment strategies more and less actively, for top and bottom distribution respectively.

6.2.2 Time series of returns

Figure 6.1: 10-year simple returns for fund category (October 2006 – October 2016)

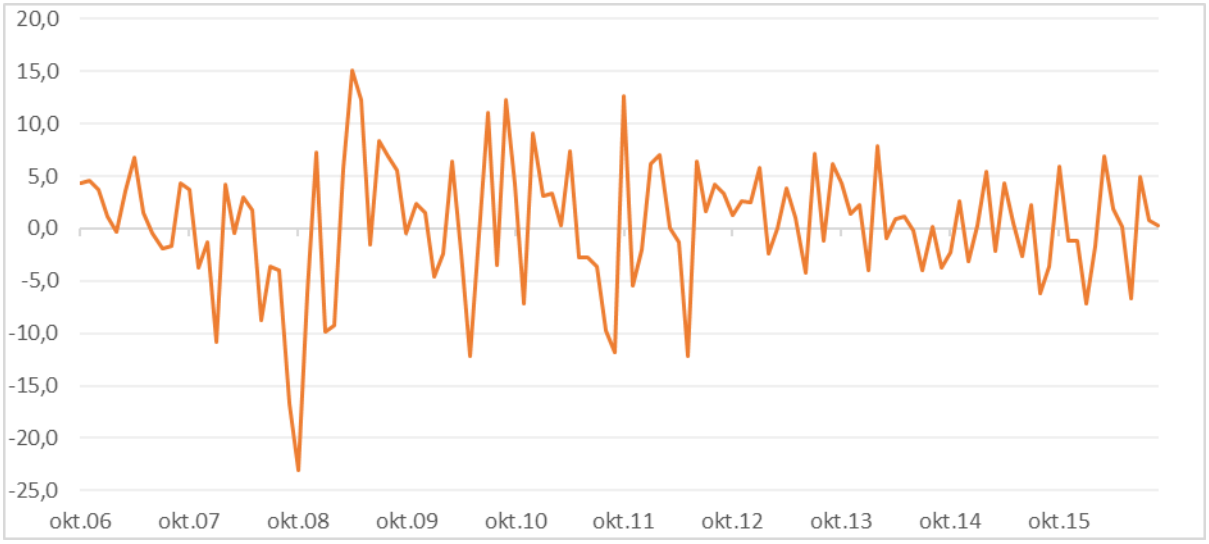


Figure 6.1 plots the ten year return series for a portfolio consisting of all funds in our fund category defined in section .1. Because of the high correlation (0.99) between market returns and fund returns, we have not plotted market returns in the same graph. The first thing we notice by looking at the plot is the huge dip in returns during the global financial crisis starting late 2007 and the jump following early 2009. Thereafter, we see a particular volatile period from 2010 and well in to 2012 that we attribute the European debt crisis. Contradicting evidence of cyclical SRI-behaviour (Johnsen and Gjølberg, 2003; Nofsinger and Varma 2014) and the presence of unique events in the global economy gives reason to shy away from periods of economic distress and rather focus on more stable time periods.

Table 6.2: Descriptiv statistics back from oct 2016 by 12, 24, 36, 60 and 120 months

Months	Mean	Std	Sharpe	Skewness	Kurtosis
12	0,25	4,61	0,05	-0,18	-0,05
24	-0,05	4,22	-0,01	0,00	-0,03
36	0,09	4,03	0,02	0,08	0,09
60	0,71	4,58	0,16	-0,06	0,64
120	0,39	6,01	0,07	-0,52	2,12

To mitigate the high variability of shorter periods (fewer observations) we require a minimum of 24 months' available returns. For the sample period we decide on 5 years making the sample period October 2011 - October 2016. We are certain that this trade-off satisfies the representativeness of a normal period and the persistence of MSR. The 5-year

monthly return series for the corresponding explanatory Fama-French factors are obtained from Kenneth French's website (Mba.tuck.dartmouth.edu., 2016). We download one dataset for three factors and one for the additional momentum factor.

6.3 Regression sample

6.3.1 **Descriptive statistics**

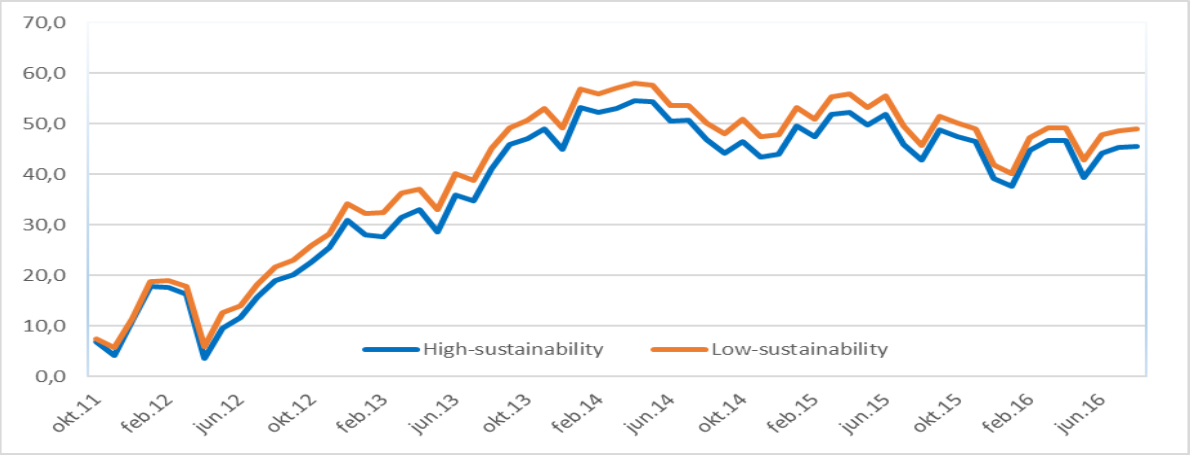
Fitting all criterion discussed so far we end up with 367 funds. The 128 high-sustainability funds yielded an annual average return of 8.53 percent. The 239 low-sustainability funds performed on average 6 basis points better, yielding an annual average return of 8.59 percent. Low-sustainability funds on average also had lower annualized variance and higher Sharpe-rate implying better risk-adjusted performance. The table below presents descriptive statistics.

Table 6.3: Descriptive statistics regression sample (October 2011 – October 2016)

	Obs	Mean	Std	Sharpe	Skewness	Kurtosis
All	367	0,71	4,58	0,16	-0,06	0,64
High	128	0,71	4,72	0,15	-0,10	0,63
Low	239	0,72	4,50	0,16	-0,04	0,63

The graph below shows again how the low-sustainability funds continuously outperforms the average high-sustainability fund over the sample period with regard to cumulative returns. The average low-sustainability fund yields a cumulative return of 49 percent while average high-sustainability fund yields 45 percent.

Figure 6.2: 5-year cumulative returns for high and low SRI-funds



6.3.2 **Management fees and transaction cost**

Gross returns are fund returns before any cost is deducted and is thus larger than net returns. Net returns deduct management, administrative, and other cost taken out of fund assets (Morningstar, 2016). This implies that the difference in gross and net is total fund expenditure and is reported in table 6.4 below. On average, high-sustainability funds have higher gross returns but lower net returns. High-sustainability funds has a monthly average fund expenditure of 17 basis points while low-sustainability funds has monthly average fund expenditure of 12 basis points.

Table 6.4: Fund expenditures

	Gross	Net	Gross-Net
High	0,88	0,71	0,17
Low	0,84	0,72	0,12
High-Low	0,05	0,01	0,04

High-sustainability funds yield a 5-year cumulative gross return of 57 percent, while low-sustainability funds yield 56 percent. This implies a 5-year cumulative fund expenditure of 12 percent for high-sustainability funds and 7 percent for low-sustainability funds. Even if this might give a clue to extraordinary cost to implementing strategies for sustainable investment we need to impose asset pricing models to account for relevant risk factors and when running regressions later in the analysis we ensure robustness by using net returns.

6.4 Survivorship Bias

Survivorship bias is a type of sampling bias that is typically addressed in relation to mutual fund studies and concerns with how omitting non-surviving funds from the sample can have implications for the results obtained (Carhart et al., 2002). Non-surviving funds refer to funds that no longer are operational, which in practice mean that they either have been merged into other funds or liquidated. A type of conditioning on survivorship typically found in empirical studies is the end-of-sample condition that only includes funds that are existing at the end of the sample period. We have constructed our regression sample on this condition, i.e. we only include funds that are still in existence as of October 2016 in addition to a requirement of minimum 24 months' returns. The problem of survivorship bias is typically tackled using a survivor-bias-free data set (e.g. Carhart, 1997) but such data sets are

obviously not suitable for our purpose due to the sustainability rating dimension. Carhart et al. (2002) find evidence that non-surviving funds underperform against surviving funds by measure of abnormal returns represented by the four-factor alpha. Non-survivors achieve a negative alpha of 0.33 percent per month, while surviving funds earn a negative 0.07 percent. Evidence is found that the bias (i.e. direct negative impact on performance) increase with the sample length. Carhart et al. (2002) also find a 7 basis point annual bias in their 1-year sample, while the 5-year sample show a 40 basis point annual bias. Looking at factor loadings in the Carhart four-factor model, Carhart et al. (2002) suggest that surviving funds load more on large cap and value stocks compared to non-surviving funds.

6.4.1 Matching characteristics

In our regression sample we have two fund categories, high-sustainability and low-sustainability. A problem occurs because Morningstar does not supply sustainability ratings for dead funds and we are thus faced with the problem of attributing the performance of dead funds to the two subcategories. We obtain data on dead funds and call this the dead fund sample. The dead fund sample, other than not having sustainability ratings, fit our fund category definitions from section 6.1.2. Since we are unable to categorize dead funds by sustainability ratings, we look at other characteristics that might suggest an asymmetrical distribution of dead funds to either subcategory. We hypothesis that dead funds are symmetrically distributed, i.e. poor performing dead funds are neither over- nor underrepresented in either category. The table below present descriptive characteristics for four categories where category sample refers to all surviving high- and low-sustainability funds.

Table 6.5: Descriptive statistics fund allocation to size, book-to-market and sector

	Large	Small	Value	Growth	Cyclical	Defensive	Sensitive
Sample	91	9	25	36	41	23	32
High	94	6	29	32	44	20	29
Low	87	13	22	39	39	24	28
Dead	85	10	31	33	42	22	32

We look at the differences in market cap exposure, in which large and small groups are defined as top 90 and bottom 10 percent, respectively. We also look at two book-to-market equity exposures, value (high book-to-market) and growth (low book-to-market. We look at three sector exposures, namely cyclical, defensive and sensitive. Descriptive statistics for our

dead fund sample in table 6.5 indicate no difference exposure to small cap relative to the surviving funds. This is in contrast with the evidence of Carhart et al. (2002) stating that surviving funds tend to load more on large cap. More importantly, there is an indicator that high-sustainability are more exposed to large cap relative to low-sustainability funds. This is in line with the findings of Bauer, Koedijk, and Otten (2005), Nofsinger and Varma (2014). Relative to the non-surviving sample the surviving sample does not load more on growth as suggested by Carhart et al. (2002). Descriptive statistics suggests that low-sustainability funds load more on growth than high-sustainability. This is in line with the findings of Nofsinger and Varma (2014), but contrast Renneboog et al. (2008) and Bauer, Koedijk, and Otten (2005). Surviving funds are according to Carhart et al. (2002) more exposed to growth stocks, but the dead fund sample fall in between the two subcategories and thus do not share any significant BE/ME characteristics with either subcategory. Compared to the surviving sample the non-survivors load a little less on defensive and a little more on cyclical. The high and low categories have distinctive differences in defensive and cyclical loadings, but the dead fund sample again falls somewhere between loadings for the high and low portfolios making the interpretation inconclusive. We find no convincing arguments against the hypothesis that neither of the two categories share more common characteristics to dead funds than the other.

6.4.2 **Inherent sampling attrition**

In the remainder of this section we address the inherent selection bias of dead fund returns making us view dead fund returns with caution. The first thing we notice by looking at table 6.6 is that dead funds historically have performed better than the other categories which seem counter-intuitive. The dead funds yield a monthly average return of 1.30 percent and a 5-year cumulative return of 43 percent. Dead funds are more volatile but also have significantly higher Sharpe-ratios than all other categories.

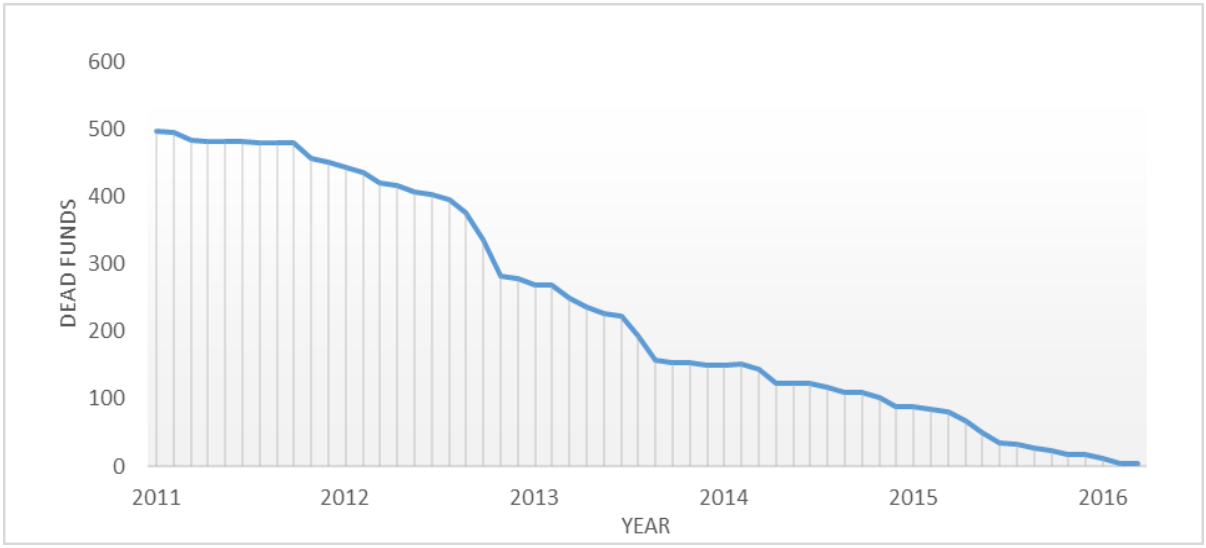
Table 6.6: 5 year descriptive statistics for comperative categories

	Obs	Mean	Std	Sharpe	5Y-Cum
Sample	367	0,71	4,58	0,16	47
High	128	0,71	4,72	0,15	45
Low	239	0,72	4,50	0,16	49
Dead	509*	1,30	5,40	0,24	43

** referes to sample attrition where in the beginning of the sample there is 509 observations*

If we look back at figure 7.2 we see that cumulative returns for the regression sample start to level out in the beginning of 2014. The returns of dead funds and regression sample is highly correlated with a coefficient of 0.988 meaning that have relatively similar return distributions. We find a maximum point of 56 percent for the full regression sample by June 2014. From this point on returns stop to have positive accumulation effect and cumulative returns drops down to 47 percent by October 2016. This indicates poor returns in the last two-three years of our sample period. We do a simple test and split the 60-month regression sample into two 30-month sub-periods. To no surprise we find significantly higher average returns in the first 30 months (1.70 percent) than in the last 30 months (-0.65 percent). It seems reasonable to say that the dead funds were weaker funds not resilient enough to survive the period of poor returns that started in the beginning of 2014 and persisted out the sample period. In figure 6.3 below we observe the fund attrition that happens as a consequence of the poor return period.

Figure 6.3: Attrition of return observations related to dead funds



The practical implication on the sample is that we have an overrepresentation in terms of returns in the beginning of the period (2011) and an underrepresentation of returns in the end (2016). This type of selection bias can according to Ang (2014) make us overestimate expected return and underestimate volatility. Statistical inference about the population mean is of special importance and is based on the sample mean that is calculated by $\bar{X} = \frac{X_1 + X_2 \ldots + X_n}{n}$. The sampling distribution of $\bar{X}$ has a mean I) E ($\bar{X}$) = μ and a standard deviation II) $\mathrm{SD}\ (\bar{X}) = \sigma/\sqrt{n}$. The population mean is expressed in terms of μ and

population standard deviation σ. The first statement shows that the distribution of $\bar{X}$ is centred at the population mean μ in the sense that expectation serves as a measure of the distribution mean. The second statement show that the standard deviation of $\bar{X}$ is equal to the population standard deviation divided by the square root of the sample size. The variability of the sample mean is governed by the population variability σ and the sample size *n*. Large variability in the population implies large variability in $\bar{X}$ making the sample information about μ less dependable. This can be countered by choosing a larger sample size *n*. Increasing the sample size will decrease the standard deviation and the distribution of $\bar{X}$ tends to become more concentrated around the population mean. In other words, as $n \to \infty$, $\bar{X} \to \mu$.

7. Research methodology

The risk and return characteristics presented in our descriptive statistics does not accurately delineate the financial performance of SRI. Fundamental principles in finance build on risk-reward relationship that state that investors only are compensated for exposure to systematic risk. In order to obtain an accurate estimate of the performance and investment style differences in mutual funds with high and low sustainability ratings we impose three different factor models: CAPM (Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black, 1972), Fama-French three-factor model (1993) and Carhart four-factor model (1997). The intuition behind factor models is that assets earn risk premiums because of their exposure to systematic risk factors (Ang, 2014). All models are estimated using ordinary least squares (OLS) methodology.

Funds are categorized into high- and low sustainability in which we construct two equal-weighted portfolios of monthly returns for the two categories. The primary objective of our study is to examine the difference in performance and investment style so we create a difference portfolio where we go long in high-sustainability funds and short in low-sustainability funds. This is practically solved by subtracting the average returns of the low-SRI portfolio from the average returns of the high-SRI portfolio. Loadings imposed by the factor models serves as evidence of differences in exposure to risk factors and abnormal return.

CAPM builds on the mean-variance efficient portfolio of Markowitz (1952) and explains the risk-reward relationship for an asset where risk is represented by the single market factor. Formally, we estimate the following CAPM-based single-factor model for both the high- and low-sustainability portfolio:

$$r_{it} - r_{ft} = \alpha_t + \beta_1(r_{mt} - r_{ft}) + \varepsilon_{it}, \quad (1)$$

where r_{it} is the return on portfolio *i* in month *t*, $\mathbf{r_{ft}}$ is the risk-free rate in month *t*, r_{mt} is the market return in month *t*, and ε_{it} is an error term. Consequently, β_1 measures the market risk exposure and α_t represents Jensen's alpha as introduced by Jensen (1968), i.e. the

monthly excess return above compensation for factor loadings.

CAPM has been subject to an extensive empirical testing regarding its practical relevance. Fama and French (2004) address the empirical evidence that invalidates CAPM's use in practise. Tests on CAPM are based on three implications arising from the market factor and the expected return relationship the model is built on. Relevant here is the linear relationship between expected returns and interrelated market factor implying that no other variable has marginal explanatory power. The hypothesis that the market factor is sufficient in explaining expected returns has been empirically rejected several times (Banz, 1981; Basu, 1983; Bhandari, 1988; Stattman, 1980; Rosenberg, Reid and Lanstein, 1985). Fama and French (1993) provide further evidence that the empirically determined variables size and book-to-market equity (BE/ME) work as additional common risk factors that is shared and undiversifiable in the stock market. The market portfolio works as a proxy for the market factor and is the value-weighted average of all the portfolios used to form the following two factors. The SMB portfolio is meant to mimic the return advantage of investing in small stocks opposed to big stocks. Small stocks are generally considered riskier than big stocks, and therefore has a higher expected return. The HML portfolio is meant to mimic the return advantage of investing in value stocks (high BE/ME) opposed to growth stocks (low BE/ME). Value stocks are considered to systematically outperform growth stocks. We estimate the three-factor model as specified in equation (2):

$$r_{it} - r_{ft} = \alpha_t + \beta_1(r_{mt} - r_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \varepsilon_{it}, \quad (2)$$

where SMB, or Small Minus Big, is the difference in return between a small cap portfolio and a large cap portfolio in month t and HML, or High Minus Low, is the difference in return in month t between a portfolio containing value stocks and one consisting of growth stocks. β_2 and β_3 are the loadings on the size and value factor, respectively.

Carhart (1997) find evidence that persistence shown in the performance of mutual funds is not explained by stock-picking skills of portfolio managers. The author argues that the predictability in mutual fund returns almost exclusively is explained by common factors in stock returns and persistent differences in fund expenditure. Jegadeesh and Titman (1993) find that strategies that involve buying stocks that has done well in the past and selling stocks that has performed poorly yield significant positive returns over a three to twelve-month period. The momentum strategy suggest that asset managers employ an investment

strategy where they methodically pick stocks on the performance of their past returns. The strategy is to buy assets that have done well (winners) during the past year and sell assets that have performed poorly (losers) in the same period. Carhart (1997) explain this return advantage not by the investment strategy indicated by Jegadeesh and Titman (1993) but rather that some mutual funds just happen to load more on past winners. Carhart (1997) findings show similar characteristics to Fama and French (1993) model and potentially a substantial explanatory effect in the WML factor which works as a proxy for the momentum factor. Consequently, we also estimate the following four-factor model for the high-SRI and low-SRI portfolio:

$$r_{it} - r_{ft} = \alpha_t + \beta_1(r_{mt} - r_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 WML_t + \varepsilon_{it}, \quad (3)$$

where WML_t, or Winners Minus Losers, is the difference in return between a portfolio of past 12 months' winners and a portfolio of past 12 months' losers in month t.

8. Empirical analysis

8.1 Main findings

Table 8.1 reports regression results obtained from our factor models. Reported are the results for portfolios comprising high-ranked and low-ranked funds, as well as the difference portfolio. The main conclusions are drawn from the multifactor models, as these tend to have greater explanatory power.

Table 8.1: Regression results

	CAPM			Fama-French three-factor			Carhart four-factor		
	High	Low	Difference	High	Low	Difference	High	Low	Difference
Alpha	-0.04 (-0.49)	0.04 (0.52)	-0.08 (-1.41)	-0.16* (-2.00)	-0.06 (-0.92)	-0.10 (1.58)	-0.13* (-1.94)	-0.07 (-1.07)	-0.06 (-0.99)
MKT	1.07*** (32.38)	1.01*** (38.42)	0.06* (1.89)	1.12*** (34.13)	1.05*** (42.41)	0.07** (2.48)	1.12*** (37.39)	1.05*** (43.70)	0.07** (2.51)
SMB				0.12* (2.32)	0.19* (1.87)	-0.07* (-1.94)	0.11** (2.33)	0.19* (1.84)	-0.07** (2.05)
HML				-0.19*** (-4.36)	-0.25*** (-7.80)	0.06 (1.18)	-0.19*** (-4.18)	-0.26*** (-6.51)	0.06 (1.30)
WML							-0.024 (-0.68)	0.008 (0.31)	-0.032 (-1.34)
Adj R^2	0.94	0.93	0.03	0.97	0.98	0.09	0.98	0.98	0.10

*Table 8.1 report the results of the estimation of Eq. (1), (2) and (3) for the 2011:10-2016:09 period. Reported are the OLS estimates for both high-SRI (High) and low-SRI (Low) portfolios. Difference is a portfolio which is constructed by subtracting the low-SRI portfolio returns from the returns of the high-SRI portfolio. The t statistics are presented in parentheses *p < 0.10, ** p < 0.05, *** p < 0.01. All regressions use Newey-West standard errors with four lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West, 1987).*

8.1.1 **Performance – Alpha**

We do not find statistical evidence that high-sustainability funds either outperform or underperform to funds with low sustainability rating after controlling for common risk factors. Alphas pertaining to the difference portfolio is negative and not statistically

significant in all three models. Looking at individual portfolios, alphas for the high-SRI portfolio are negative and statistically significant in both multivariate regressions. This indicates that high-sustainability funds underperform relative to the factor benchmark which are the explanatory Fama-French return data in the regression, with a monthly 16 basis points (Fama-French three-factor model) and 13 basis points (Carhart four-factor model). Alphas for the low-SRI portfolio are relatively small, negative although not statistically significant in all three models. Our findings are not surprising given the descriptive statistics displayed in table 6.3 which show only small differences in the reward-to-risk (Sharpe) ratio indicating that neither portfolio would outperform the other. Our findings are generally consistent with previous empirical studies which largely concludes that there is neither a gain nor a loss associated with SRI contra conventional investing (Statman, 2000; Johnsen and Gjølberg, 2003; Bauer, Koedijk and Otten, 2005).

8.1.2 **Investment style – Market loading**

Regression results provide evidence that high-sustainability funds are more exposed to market risk relative to low-sustainability funds. Market loadings are positive and statistically significant for the difference portfolio in all models. In CAPM the difference portfolio has a market loading of 6 basis points while in the Fama-French three-factor model and Carhart four-factor model the difference portfolios both show a market loading of 7 basis points. The difference portfolio goes up and down by the mentioned magnitude relative to a one percentage point change in how the explanatory Fama-French market portfolio proxy move. If the market portfolio in a given period yield a return of 1 percent, then market exposure in the difference portfolio would contribute with a 1.07 percent return according to both multivariate models.. Regarding individual portfolios, loadings and t-statistics for the market factor imply that high-SRI and low-SRI portfolios both are significantly exposed to the market factor. Except for the low-SRI portfolio in CAPM, both portfolios have loadings greater than 1 in all models which indicate greater risk than the market portfolio. Our findings are consistent with those of Nofsinger and Varma (2014) who find a positive difference in market loading between SRI's and conventional funds. Our findings where indicated by the descriptive statistics presented in table 6.5 which show that the high-SRI portfolio hold relatively more of cyclical stocks and less of defensive stocks compared to the low-SRI portfolio Our findings contrast Renneboog et al. (2008) and Bauer et al. (2005) who find a negative difference in market loading for SRI's and conventional funds.

8.1.3 **Investment style – Size loading**

We find evidence that high-sustainability funds invest relatively less in small capitalization stocks than low-sustainability funds and thus are less exposed to risk associated with small-cap investing. Loadings on the size factor are negative and statistically significant for the difference portfolio in both multivariate models. In both the Fama-French three-factor model and the Carhart four-factor model the difference portfolio has a size loading of 7 basis points. The difference portfolio moves with this magnitude relative to a one percentage point change in how the explanatory Fama-French size proxy move. If the Fama-French factor benchmark in a given period yield a 1 percent return strictly due to size-investing the difference portfolio would yield 0.93 percent according to both multivariate models. Regression results with respect to the individual portfolios demonstrate statistically significant positive size loadings for both the high- and low-SRI portfolios. Descriptive statistics in table 6.5 show that low-SRI funds load more than double on small cap relative to high-SRI. Our findings are consistent with those of Bauer et al. (2005) who find a positive difference in size loading of 7 basis points in an international perspective between SRI's and conventional funds. Renneboog et al. (2008) also document a tilting toward small-cap stocks where UK SRI funds load 22 basis points more on size than conventional UK funds.

8.1.4 **Investment style – Value loading**

We do not find statistical evidence that high-sustainability funds are more or less exposed to risk associated with value-investing relative to low-sustainability funds. Loadings on value are positive but not significant from a statistical perspective for the difference portfolio in both multivariate models. As for individual portfolios, we observe statistically significant negative loadings on value in both multivariate models. Thus, indicating a tilting towards growth stocks for both high- and low-SRI portfolios relative to the Fama-French explanatory factor benchmark. When the factor benchmark yield a 1 percent return strictly due to the value-strategy, the high-SRI portfolio yield a 0.82 percent return while the low-SRI portfolio yield a 0.78 percent return in the Fama-French three-factor model. Individual coefficients are practically identical for both multivariate models. Descriptive statistics presented in table 6.5 show that both high- and low-SRI portfolios generally are tilted towards growth stocks, although the high-SRI portfolio load more on value and less on growth relative to the low-SRI portfolio. Our findings are in line with Renneboog et al. (2008) who finds no significant difference in value loading between SRI and conventional investments.

8.1.5 **Investment style – Momentum loading**

We find no statistical evidence of persistence in returns. Loadings on momentum are small, negative and statistically insignificant for all three portfolios.

8.2 Robustness checks

Throughout the paper, we argue that the strength of our study is that we use standardized objective measures of sustainability to evaluate financial performance. In order to check if this argument holds valid we perform additional regressions where we apply two different sustainability rankings in accordance with Morningstar Sustainability Rating. First, we wish to control to see if using more extreme top and bottom ranking thresholds relative to we base our main results on, are going to characterise the results even further in terms of statistical significance and coefficient magnitude. If this holds true, we can with higher certainty say that the sustainability component is attributing any difference in financial performance or risk difference. Thus, clarifying a causal relationship between sustainability and performance. We therefore run regression on the ranking thresholds stated below (regressions results are presented in appendix 1):

1) Conservative threshold: Regression where the high-SRI portfolio is defined as those with the top 5 percent Morningstar Sustainability Rating and the low-SRI portfolio with the bottom 5 percent rating.

The second robustness check is motivated by the same argument stated above, but by using the opposite approach to highlight the importance of non-dichotomous ranking. By using wider thresholds then our original top and bottom 10 percent sustainability we wish to control to see if a sustainability categorizing based on wider thresholds in turn will make results less characterized in terms of statistical significance and coefficient magnitude. Thus, justifying our methodology in terms of using the Morningstar Sustainability Rating. We therefore run the regression on the ranking thresholds stated below (regression results presented in appendix 2):

2) Wider threshold: Regression where the high-SRI portfolio is defined as those with the top 32.5 percent Morningstar Sustainability Rating and the low-SRI portfolio

with the bottom 32.5 percent rating. This sorting only leaves out MSR funds categorized as average, which accounts for 35 percent of funds.

Generally, alphas pertaining the difference portfolios remain statistically insignificant for all models in line with main results. Thus, more and less extreme sorting thresholds do not show sign of abnormal returns, which in turn provides further confidence that there is no causal relationship between sustainability and financial performance. As for market loading, the 5 percent thresholds show similar patterns as our main results, although loadings are of greater magnitude. The 32.5 percent threshold yields insignificant loadings which is to be expected from the looser thresholds because of less characterised sorting differences. For size loading, the 5 percent threshold again provide the same results as our main findings, yielding loadings of greater magnitude characterising our findings additionally. Again, for the 32.5 percent threshold results are not as conclusive suggesting that found characteristics are too similar to yield any results on difference portfolios. As for value and momentum factors, results remain similar and negligible for all specifications.

9. Concluding remarks

In this study, we expand the previous literature on SRI performance by investigating the return and investment style differences of European open-end mutual funds with high and low sustainability rating, respectively. By the use of three factor models, we account for differences in common risk factors. We find no statistical evidence to support that there exists a risk-adjusted difference in performance between high- and low-sustainability funds. We do however find that high-sustainability funds are more exposed to market and less exposed to small-cap risk relative to low-sustainability funds. The results are robust for different sustainability thresholds, implying a more causal relationship between sustainability and investment style characteristics and financial performance. One limitation to our study is the cross-sectional nature of the sustainability rating. To further ensure causality one would require a time series nature in measures of both financial and sustainable performance. This would also implicitly mitigate any survivorship bias.

.Our findings of neutral performance implies that there is no evidence of an additional cost related to investing in sustainability. Thus, investors can invest in sustainable and responsible mutual funds to achieve utility from sustainability efforts without any financial implications. This is relevant for investors that in addition to financial performance also derive utility from the socially responsible attributes (Bollen, 2007; Statman, 2014). Our findings are relevant for the world's biggest market for sustainable investing, namely Europe.

References

Ang, A. (2014). *Asset management*. 1st ed. New York: Oxford University Press, pp. 83, 213-238.

Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), pp.3-18.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks. *Journal of Financial Economics*, 12(1), pp.129-156.

Bauer, R., Koedijk, K. and Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7), pp.1751-1767.

Bhandari, L. (1988). Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. *The Journal of Finance*, 43(2), pp.507-528.

Black, F. (1972). Capital Market Equilibrium with Restricted Borrowing. *The Journal of Business*, 45(3), pp.444-455.

Bollen, N. (2007). Mutual Fund Attributes and Investor Behavior. *Journal of Financial and Quantitative Analysis*, 42(03), pp.683-708.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), pp.57-82.

Carhart, M., Carpenter, J., Lynch, A. and Musto, D. (2002). Mutual Fund Survivorship. *Review of Financial Studies*, 15(5), pp.1439-1463.

Cook, J., Oreskes, N., Doran, P., Anderegg, W., Verheggen, B., Maibach, E., Carlton, J., Lewandowsky, S., Skuce, A., Green, S., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R. and Rice, K. (2016). Consensus on consensus: a synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, 11(4), pp.1-7.

Core, J., Guay, W. and Rusticus, T. (2006). Does Weak Governance Cause Weak Stock Returns? An Examination of Firm Operating Performance and Investors' Expectations. *The Journal of Finance*, 61(2), pp.655-687.

Cremers, M. and Ferrell, A. (2014). Thirty Years of Shareholder Rights and Firm Value. *The Journal of Finance*, 69(3), pp.1167-1196.

Deng, X., Kang, J. and Low, B. (2013). Corporate social responsibility and stakeholder value maximization: Evidence from mergers. *Journal of Financial Economics*, 110(1), pp.87-109.

Di Giuli, A. and Kostovetsky, L. (2014). Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of Financial Economics*, 111(1), pp.158-180.

Døskeland, T. (2016). Lecture FIE426, NHH, *Investeringsprossesen.*

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3), pp.621-640.

Ehrlich, P. and Ehrlich, A. (1990). *The population explosion*. 1st ed. New York: Simon and Schuster, pp.38-39.

Eurosif, (2016). *European SRI Study*. [online] Available at: http://bit.ly/2guinxX [Accessed 11 Dec. 2016].

Fama, E. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), pp.383-417.

Fama, E. and French, K. (2004). The Capital Asset Pricing Model: Theory and Evidence. *Journal of Economic Perspectives*, 18(3), pp.25-46.

Hong, H. and Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1), pp.15-36.

Houghton, J. (2004). *Global warming*. 1st ed. Cambridge, UK: Cambridge University, p.326.

IPPC (Intergovernmental Panel on Climate Change). 2014. Climate change 2014: Mitigation of Climate Change. Working Group III contribution to the IPCC Fifth Assessment Report. Cambridge, United Kingdom: Cambridge University Press

J. Derwall, K. Koedijk, J. Ter Horst A tale of values-driven and profit-seeking social investors. *Journal of Banking and Finance*, 35 (8) (2011), pp. 2137–2147

Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1), pp.65-91.

Jensen, M. (1968). THE PERFORMANCE OF MUTUAL FUNDS IN THE PERIOD 1945-1964. *The Journal of Finance*, 23(2), pp.389-416.

Johnsen, T. and Gjølberg, O. (2003). *Evaluering av etisk forvaltning: metode, resultat og kostnader*. [online] pp.171-217. Available at: https://www.regjeringen.no/no/dokumenter/nou-2003-22/id118914/ [Accessed 11 Dec. 2016].

Johnsen, T. and Gjølberg, O. (2007). *En oppdatert empirisk analyse av erfaringer 2003*-2007. [online] pp.171-217. Available at: https://www.regjeringen.no/globalassets/upload/fin/.../analyse_etisk_forvaltning.pdf Accessed 11 Dec. 2016].

Jørgensen, S. and Tynes Pedersen, L. (2015). *Responsible and profitable*. 1st ed. Oslo: Cappelen Damm Akademisk, pp.24-27, 46.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), pp.263-291.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1), p.13.

Mba.tuck.dartmouth.edu. (2016). Kenneth R. French - Data Library. [online] Available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research [Accessed 9 Dec. 2016].

Morgan Stanley Institute for Sustainable Investing, (2015). *Sustainable Signals*. [online] pp.1-8. Available at: http://www.morganstanley.com/sustainableinvesting/pdf/Sustainable_Signals.pdf [Accessed 13 Dec. 2016].

Morningstar.com, (2016). *Equity Class Classification*. [online] Available at: http://corporate.morningstar.com/EquityClassifications/ [Accessed 11 Dec. 2016].

Morningstar.no. (2016). *Morningstar Sustainability Rating*. [online] Available at: http://www.morningstar.no/no/news/148087/morningstar-sustainability-rating.aspx [Accessed 19 Dec. 2016].

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica*, 34(4), p.768.

Newey, W. and West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), pp.703-708.

Nofsinger, A. and Varma, V. (2014). Socially responsible funds and market crises. *Journal of Banking & Finance*, 48, pp.180-193.

Pedersen, L. (2015). *Efficiently inefficient*. 1st ed. New Jersey: Princeton University Press, pp.21-22.

People.stern.nyu.edu. (2016). Useful Data Sets. [online] Available at: http://people.stern.nyu.edu/adamodar/New_Home_Page/datacurrent.html#multiples [Accessed 9 Dec. 2016].

Porter, M. and Linde, C. (1995). Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, 9(4), pp.97-118.

Renneboog, L., Ter Horst, J. and Zhang, C. (2008). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14(3), pp.302-322.

Rosenberg, B., Reid, K. and Lanstein, R. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management*, 11(3), pp.9-16.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3), pp.425-442.

Statman, M. (2000). Socially Responsible Mutual Funds. *Financial Analysts Journal*, 56(3), pp.30-39.

Statman, M. (2014). Behavioral finance: Finance with normal people. *Borsa Istanbul Review*, 14(2), pp.65-73.

Statman, M. and Glushkov, D. (2009). The Wages of Social Responsibility. *Financial Analysts Journal*, 65(4), pp.33-46.

Stattman, D. (1980). Book values and stock returns. The Chiacgo MBA*: A Journal of Selected Papers*, 4, pp.25-45.

Strategy and Society: The link between competitive advantage and corporate social responsibility. (2006). *Harvard Business Review*, (84:12), pp.78-92.

Sustainalytics, (2016). *Research Methodology*. [online] Available at: http://marketing.sustainalytics.com/acton/formfd/5105/004e:d-0002 [Accessed 19 Dec. 2016].

The US SIF Foundation, (2016). *Sustainable and responsible investing trend*. [online] Available at: http://ussifnewsite.membershipsoftware.org/files/SIF_Trends_16_Executive_Summary(1).pdf [Accessed 11 Dec. 2016].

Un.org. (2016). *Population | UN DESA | United Nations Department of Economic and Social Affairs*. [online] Available at: https://www.un.org/development/desa/en/key-issues/population.html [Accessed 17 Dec. 2016].

UNglobalcompact.org. (2016). *Social Sustainability | UN Global Compact*. [online] Available at: https://www.unglobalcompact.org/what-is-gc/our-work/social [Accessed 17 Dec. 2016].

United Nations, (1987) *Our Common Future - Brundtland Report*. Oxford University Press, p. 204.

Unpri.org. (2016). *UN Principles for Responsible Investment*. [online] Available at: https://www.unpri.org/about [Accessed 19 Dec. 2016].

Vff.no. (2016). *Bransjestandarder - VFF*. [online] Available at: http://vff.no/bransjestandarder#informasjon-og-klassifisering-av-aksje-og-kombinasjonsfond [Accessed 13 Dec. 2016].

Wimmer, M. (2012). ESG-persistence in Socially Responsible Mutual Funds. *Journal of Management and Sustainability*, 3(1), pp.9-15.

Appendix

Appendix 1: Robustness check: 5 percent threshold

Regression results – 5 percent sustainability rating threshold

	CAPM			Fama-French 3-factor			Carhart 4-factor		
	High	Low	Difference	High	Low	Difference	High	Low	Difference
Alpha	-0.04	-0.05	0.01	-0.17*	-0.09	-0.07	-0.07	-0.08	0.01
	(-0.37)	(-0.57)	(0.09)	(-1.68)	(-1.16)	(-0.83)	(-0.86)	(-1.19)	(0.08)
Market	1.11***	1.05***	0.06**	1.17***	1.07***	0.10**	1.18***	1.07***	0.10***
	(24.11)	(36.95)	(2.05)	(26.98)	(35.71)	(2.58)	(30.83)	(39.64)	(2.97)
Size				0.13***	0.21	-0.08**	0.13***	0.21*	-0.08**
				(2.93)	(1.63)	(-2.46)	(2.87)	(1.79)	(-2.24)
Value				-0.18***	-0.22**	0.04*	-0.19***	-0.22***	0.04
				(-3.23)	(-2.58)	(1.87)	(-3.49)	(-2.71)	(1.51)
Momentum							-0.08	-0.01	-0.07
							(-1.49)	(-0.29)	(-1.57)
R^2	0.96	0.98	0.07	0.97	0.98	0.21	0.97	0.98	0.26
Adjusted R^2	0.96	0.98	0.06	0.97	0.98	0.17	0.97	0.98	0.20

*The table reports the results of the estimation of Eq. (1), (2) and (3) for the 2011:10-2016:09 period. Reported are the OLS estimates for both high-SRI (High) and low-SRI (Low) portfolios. Difference is a portfolio which is constructed by subtracting the low-SRI portfolio returns from the returns of the high-SRI portfolio. The t statistics are presented in parentheses *p < 0.10, ** p < 0.05, *** p < 0.01. All regressions use Newey-West standard errors with four lags, where the number of lags is determined by 4(n/100)$^{2/9}$ (Newey and West, 1987).*

Appendix 2: Robustness check: 32.5 percent threshold

Regression results - 32.5 percent sustainability rating threshold

	CAPM			Fama-French 3-factor			Carhart 4-factor		
	High	Low	Difference	High	Low	Difference	High	Low	Difference
Alpha	0.02 (0.26)	0.01 (0.26)	0.01 (0.13)	-0.09 (-1.65)	-0.07 (-1.31)	-0.03 (-0.73)	-0.09 (-1.64)	-0.06 (-1.25)	-0.03 (-0.47)
Market	1.06*** (26.73)	1.03*** (39.66)	0.02 (1.05)	1.07*** (27.36)	1.07*** (40.64)	0.01 (0.31)	1.07*** (30.14)	1.06*** (39.64)	0.01 (0.40)
Size				0.06 (1.47)	0.07 (1.72)	-0.02 (-1.63)	0.05 (1.49)	0.07* (1.83)	-0.02* (-1.73)
Value				-0.13*** (-4.57)	-0.18*** (-5.87)	0.05 (1.56)	-0.14*** (-4.10)	-0.18*** (-4.86)	0.05* (1.97)
Momentum							-0.01 (-0.16)	0.00 (-0.08)	-0.01 (-0.19)
R^2	0.98	0.99	0.08	0.98	0.99	0.22	0.98	0.99	0.24
Adjusted R^2	0.98	0.99	0.07	0.98	0.99	0.17	0.98	0.99	0.21

*The table reports the results of the estimation of Eq. (1), (2) and (3) for the 2011:10-2016:09 period. Reported are the OLS estimates for both high-SRI (High) and low-SRI (Low) portfolios. Difference is a portfolio which is constructed by subtracting the low-SRI portfolio returns from the returns of the high-SRI portfolio. The t statistics are presented in parentheses *p < 0.10, ** p < 0.05, *** p < 0.01. All regressions use Newey-West standard errors with four lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West, 1987).*