



Norwegian University
of Life Sciences

Master's Thesis 2019 30 ECTS

School of Economics and Business, NMBU

Sustainable Investing

Is there a relationship between ESG ratings and fund performance?

Julian Taghawi Moussawi

Master of Science in Business Administration

ABSTRACT

This paper studies the performance of 146 mutual funds with a Norwegian International Securities Identification Number (ISIN). Dividing the sample into quintiles based on a variety of metrics within Morningstar Sustainability Rating, I find no evidence of a marked-based outperformance by the top or bottom quintile funds. This would indicate that there currently exists no financial benefits or drawbacks to investing in high (or low) ESG-rated funds. However, there is a recurring notion of a geographical bias in the distribution of sustainability ratings. When isolating European-categorized funds (n=67) in the data for a more homogenous investment universe, several alphas are significant at the 10% level. The results show that the superior sustainability metric in the study; the Historical Portfolio Sustainability Score provides a monthly alpha of 0,4% in the top rating quintile and a 0,3% monthly excess alpha over the bottom quintile. In accordance with previous research, superior governance performance is found to influence risk-adjusted returns in a positive, statistically significant manner. Furthermore, I find evidence of an existing ESG-momentum effect in the Norwegian mutual fund market. Performance gains from increasing ESG-scores could indicate an existing reward connected to buying low sustainability funds and investing efforts into improving the funds' sustainability through being an active shareholder.

PREFACE

This thesis represents the end of a two-year master's degree program in Business Administration at the Norwegian University of Life Sciences. Being able to write about something as interesting as sustainable investing made the process fun and truly educational.

Rightfully, there are several people who needs thanking for the final product. First, a big thank you to my supervisor Ole Gjølberg, who through persistent suggestions provided me with a topic that finally caught my interest. I am appreciative of the concise and insightful feedback as well as the support throughout the process.

Secondly, thank you to Christopher Greiner and Bjørn Leander at Morningstar who were very accommodating and provided the data necessary to conduct this study.

Finally, thank you to my family and friends who have been there for me for the challenging past couple of years. You know who you are.

Ås, 2019

Julian Taghawi Moussawi

CONTENTS

1. Introduction	5
1.1 Background, thesis objective and research value	6
2. Defining sustainability – AUM and terminology	9
2.1 Sustainability related AUM	9
2.2 Sustainability related terminology	10
3. The Morningstar Sustainability Rating. An overview	14
3.1 The Morningstar- and Sustainalytics collaboration	14
3.2 Rating methodology	15
3.3 Historical Portfolio Sustainability Score	19
3.4 Environmental (E), Social (S) and Governance (G) scores	19
4. Literature review	21
4.1 Sustainability and performance	21
4.2 Meta studies on sustainability and performance	24
4.3 Morningstar Sustainability Rating and mutual fund performance	25
4.4 Controversies, controversy scores and fund evaluation	26
4.5 Critique of ESG ratings as objective measures	27
5. Theoretical framework	31
5.1 Sustainability and asset pricing	31
5.2 Performance measures and systematic risk factors	33
5.3 Cross-sectional analysis	36
6. Data	37
6.1 Data collection - describing the raw data	37
6.2 Filters on the dataset and fund categories	38
6.3 Sustainability metrics in the analysis	39
6.4 Persistence in the data	41
6.5 Descriptive statistics	44
7. Research methodology, econometrics, results and analysis	54
7.1 Fama-French three-factor regression	54
7.2 ESG-momentum	61
7.3 Cross-sectional analysis	62
8. Conclusion	68
9. Reference list	70
10. Appendix	75

List of tables

Table 3.1. Sustainalytics' company controversy score

Table 3.2. Morningstar Sustainability Rating distribution

Table 4.1. Results from Clark et al (2015) of ESG studies and market outperformance

Table 4.2. Average ESG Score by market cap

Table 6.1. Overview of fund allocation by category

Table 6.2. Descriptive statistics of top and bottom quintile sustainability portfolios

Table 6.3. Percentage of allocation in large-cap categorized funds and in Europe

Table 6.4. ESG momentum portfolio performance and sustainability statistics

Table 7.1. Fama-French three factor regression results.

Table 7.2. Fama-French three-factor regression results: European categories

Table 7.3. Fama-French regression results: ESG-momentum

Table 7.4. Cross-sectional regression results: Average returns with respect to beta values.

Table 7.5. Cross-sectional regression results: Average returns with respect to beta and HPSS.

Table A1. Portfolio allocation by fund category.

Table A2. Descriptive statistics of sustainability metrics

Table A3. Cross-sectional regression results: Sub-sample average returns with respect to beta and HPSS.

Table A4. Cross-sectional regression results: Average returns with respect to beta, HPSS, geographical dummy and top and bottom sustainability dummies.

Table A5. Cross-sectional regression results: Average returns with respect to betas and ESG-momentum.

List of figures

Figure 4.1. Renaissance Capital ESG score vs GDP per capita

Figure 5.1. MSCI top vs. bottom quintile idiosyncratic risk incidents 2007-2014.

Figure 6.1. Wimmer (2013) Persistence of ESG ratings

Figure 6.2. Persistence of ESG ratings with constant portfolio holdings

Figure 6.3. 5-Year cumulative returns for OSEFX and the full, equally weighted 146-fund sample.

Figure 6.4. Standard deviation to beta relationship

Figure 6.5. Top and bottom quintile HPSS monthly returns.

Figure 6.6. Top and bottom quintile HPSS annualized standard deviations.

Figure 6.7. Top and bottom quintile HPSS portfolio beta values.

Figure 7.1. Binomial distribution of European-categorized funds (46% of sample) in the High HPSS portfolio.

1. Introduction

The demand for “sustainability” as a part of investment strategies is rapidly increasing among financial institutions and private investors alike. Sustainable investments are reported to account for 26% of all assets under management (AUM) on a global scale, amounting to \$23 trillion, as European markets lead the way with roughly half of AUM considering sustainability criteria (GSIA, 2016).

With the intention to meet the increased market demand, funds with a self-proclaimed sustainability mandate have experienced rapid growth. These funds tend to carry costs disproportionate to the mutual fund market otherwise. Furthermore, despite the recent increase, these funds only consist of roughly 2% of the mutual fund market (Forbes, 2017). The contents of these funds are vastly diverse and generally exclude industries based on ethical considerations. Such limits on the investment universe are typically associated with an unoptimized portfolio.

By broadening the meaning of “sustainable investments,” new concepts and approaches have blossomed, amongst which was the UN-backed introduction of “ESG.” Moving away from ethical-based screening, environmental (E), social (S) and governance (G) factors are presumed to have financial relevance. Through industry-peer comparison, companies are ranked and rated based on their ability to sustain in the market. Applying this approach, a market-wide underestimation of long-term loss associated with idiosyncratic risk incidents (Nofsinger and Varma, 2014; MSCI, 2018) could imply reduced risk (at no cost) by primarily including high-scoring ESG assets. I elaborate on this in chapter 5.

Accommodating the demand, Morningstar began to offer sustainability ratings for a broad portfolio of mutual funds in 2016 with the launch of Morningstar Sustainability Rating. Being able to objectively assess sustainability of investment options, investors got newfound prospects of integrating sustainability in their portfolios.

In this paper, I cover parts of the preceding literature, provide opinions and perform statistical analysis on the topics highlighted above. What exactly is sustainable investing and ESG, how can it be integrated in to a diversified portfolio and are there any financial benefits or drawbacks to doing so? These are some of the questions I attempt to answer in this paper. My goal is to provide the reader with insight to the (sometimes confusing) subject of sustainable investing and how it scales with mutual fund performance.

1.1 Background, thesis objective and research value

In the world of finance, environmental awareness and social responsibility are hardly new concepts. Terms such as socially responsible investing (SRI) have been employed in investment strategies for decades. The world's largest sovereign wealth fund, namely The Government Pension Fund Global (The Oil Fund) values SRI criteria as they exclude industries such as tobacco and firearms in their investment strategy. Furthermore, coal companies are largely excluded or placed under observation following the product-based coal criteria¹ established by the Norwegian Ministry of Finance in 2014 (NBIM, 2019). Other examples of industries which are often subjected to negative screening as such are the oil, alcohol, gambling, nuclear power and logging and mining industries. However, negative screening comes at a cost. Hong and Kacperczyk (2007) and Dimson et al (2015) find that investing in "sin" stocks historically outperforms the market. While ethical considerations is a noble practice, negative screening based on personal values can simply allow for other investors to exploit created gaps in otherwise efficient markets. This hypothesis finds support through economic theory and academic literature, though opinions are not remotely unanimous.

In the middle of the 2000s, a new concept, ESG (environmental, social, governance) rose to the surface within the subject of sustainable investing. The now better-known term was established through a report published the United Nations Global Compact², accompanied by the Swiss Federal Department of Foreign Affairs. "Who Cares Wins – Connecting Financial Markets to a Changing World" (Knoepfel, 2004) concludes with conviction that environmental, social and governance factors are essential in order to create a sustainable, resilient investment market. Rather than being predominantly motivated by ethics as often the case with SRI, ESG factors were assumed to have financial relevance. The report had fundamental importance for the later creation of the UN-backed Principles for Responsible Investment (PRI)³.

In the report, the authors recommend that investors, from asset managers to security brokerage houses improve the integration of ESG-factors in their investment analysis. ESG-factors mentioned in the report include but are not limited to:

¹ Companies basing more than 30% of their income or business otherwise on thermal coal can be excluded.

² UN initiative supporting ESG practices.

³ UN supported network of investors working towards the incorporation of ESG-factors in international markets. PRI launched in 2006 and has since then gathered 1800 signatories worldwide, accounting for more than \$70 trillion AUM (unpri, 2019).

E: Sensitivity to regulations and efforts in reducing negative impact.

S: Workplace health and safety, community relations and human rights policies.

G: Board structure and accountability, accounting practices, corruption and bribes.

Common issues for all ESG-factors include reputational risks resulting from mismanagement.

Increased focus on ESG resulted in an increase of positive screening approaches. Through assigning a screening process, investors seek excess risk-adjusted returns by reducing the risk caused by the aforementioned factors. The strategy focuses on long-term profits earned by limiting exposure to the higher tail-risk parts of a portfolio, which are often underestimated by the market (Nofsinger and Varma, 2014).

UN's report highlighted challenges hindering integration of ESG-factors in investments. In short, challenges relate to disclosure of information, defining ESG issues and lack of competence. In order to implement ESG effectively, a degree of standardization is required. Companies need to disclose information regarding their ESG-policies such that a third party can conduct a ranking based on set criteria. While there is still no standard definition of sustainability, several agencies have attempted to accommodate for this demand. The first (at portfolio level), and perhaps the most comprehensive of which was Morningstar⁴ with the development of the Morningstar Sustainability Rating (MSR), through a partnership with Sustainalytics⁵.

With MSR, Morningstar quantified sustainability for mutual funds, contributing to both investment management and ESG-research. By providing (partly) publicly available sustainability scores, investors can objectively assess the sustainability of their investment options. Arguably more importantly, one can analyze the performance of managed portfolios at differing sustainability levels.

With the rapid growth of the market for sustainable investing, there exists extensive research on the subject. Most research is based on negative screens, corresponding with the scope of this approach, as it historically is the most dominant in terms of AUM. There is also research available on a variety of other screening characteristics, including "best-in-class" approaches, like in this paper. However, a common factor in most of these studies is the lack of quantifiable sustainability. MSR was the first methodology which ranked and rated mutual

⁴ Investment research and management company. More information in chapter 3.

⁵ Leading provider of ESG research. More information in chapter 3.

funds' sustainability. With its introduction in March 2016, MSR is still relatively new to the market and the research on its effects is limited.

The current studies on MSR include effects on fund flows and on risk-adjusted returns in the US and European equity mutual fund market. To provide comparative value, I research the relationship between MSR, and the risk-adjusted returns of mutual funds domiciled in Norway. To the best of my knowledge, current studies have limited the analysis to effects of the superior metrics within the MSR, i.e. the Portfolio ESG Score or the Portfolio Sustainability Score. In hopes of providing some absolute value I analyze the effects of sub-metrics within the MSR, including controversy scores and scores for each of the ESG-pillars. Another deviation I make from previous research is basing overall sustainability categories on the newly implemented Historical Portfolio Sustainability Score. Furthermore, I study the possibility of an "ESG momentum"-effect between an increasing sustainability score and risk-adjusted returns.

My underlying hypothesis is that companies at the peak of industry-specific sustainability are associated with increased performance. The basis of this hypothesis is found in the increased tail-risk connected to low ESG-rated funds accompanied by a market-wide underestimation of this downside risk. I outline this further in chapter 5.

Through this study, I aim to contribute to the literature regarding performance of sustainable investing, or more specifically ESG investing. I wish to provide investors with a better understanding of what ESG is and how it impacts fund performance. In a growing market, this could be of interest to any investors who wish to stay relevant in times to come.

2. Defining sustainability - AUM and terminology

As mentioned initially, sustainable investments are reported to account for 26% of global AUM. Given how “sustainable investments” or “sustainability” is not well defined, what does this figure mean? Here, I will discuss global sustainability numbers as presented by the large organizations, such as GSIA, Eurosif and US SIF as well as outline some of the issues regarding these figures. I also define how the terminology is to be interpreted in this paper.

2.1 Sustainability related AUM

The Global Sustainable Investment Alliance (GSIA) is an international collaboration between the seven largest membership-based sustainable investment organizations (Eurosif, 2019). Their biennial report is the foundation of discussion within the topic of investments based on ethical or sustainable issues. The most recent available review is from 2016. Figures presented show an increase in sustainability related AUM of 25% since 2014, from \$18,2 trillion to \$22,9 trillion. Of the global figure, \$12,0 trillion and \$8,7 trillion are invested in Europe and the United States respectively, totaling 91% of all AUM invested in so-called sustainability. Since then, The Forum for Sustainable and Responsible Investment (US SIF) published an updated review, reporting a dramatic increase to \$12,0 trillion (37,9% increase) in the United States (US SIF, 2018). During the same period, AUM for all assets “only” increased by 15,6%. The growth is largely attributed to asset managers considering ESG criteria, covering \$11,6 trillion in assets. The European Sustainable Investment Forum (Eurosif) reported increased sustainability numbers for Europe as well, though in overlapping, strategy-specific terms, i.e. not applicable for comparison.

While the figures presented above seem impressive, it is my opinion that the specifics are unimportant in the current climate. There is currently no standard protocol for reporting sustainability, no standard protocol for measuring sustainability, no standard way of defining sustainability, and the list goes on. This is reinforced by the GSIA report, declaring that the aggregated figure referred to as “sustainable investing” accounts for ESG concerns “[...] without making judgement about the quality or depth of the process applied.” Data collection differs between regions and relies heavily on questionnaires. Additionally, as reporting methods can change on a year-to-year basis, reported numbers can fluctuate heavily without being representative for previous years’ methodology. A clear example of this is found in the

2016 GSIA report, where Japan showed an increase of 6690% in sustainability-assets from 2014. Mostly, this can be attributed to changes in reporting practices.

Rather than the specifics, the take-home message is recognizing the increased awareness of, and interest in sustainability expressed by investors. That being said, more and more corporations adhere to measures and initiatives such as GRI-standards⁶ and IIRC⁷, so the degree of standardization is ever increasing.

2.2 Sustainability related terminology

While studying the topic in preparation for this thesis, it became clear that the terminology in the universe of sustainable and ethical investing lacks in universal definition. Terms are profoundly overlapping, numbers are hard to comprehend, and the wording is often unintuitive. Before delving in to the subject further, I wish to dedicate a chapter to the semantics, how terms are used in the literature and how they are to be interpreted in this thesis. My goal is not to create universal definitions, but simply to highlight the issue and define the concepts as used in this paper.

Establishing a hypernym

Because sustainable and/or ethical investing has no definite standardization, the terminology is subjectively interpreted. This often leads to confusion and misinterpretation. While researching, I have mostly come across three umbrella terms, or hypernyms used to describe all investments with some degree of ethics or sustainability in mind: “ESG”, “SRI” and “sustainable investing”. Business Insider (2018) claims that \$23 trillion AUM are “earmarked with an **ESG** mandate”. Bloomberg (2017) makes a reference to the same data but declares it “global **sustainable investments**”, while J.P. Morgan (2018) announces a \$23 trillion “global **SRI** market”.

In GSIA’s 2016 review, it is read, “Globally, there are now \$22.89 trillion of assets being professionally managed under **responsible investment** strategies [...] Clearly, **sustainable investing** constitutes a major force across global financial markets.” Again, supporting the

⁶ Global Reporting Initiative, founded in 1997, created to empower decisions that create social, environmental and economic benefits for everyone (GRI, 2019).

⁷ The International Integrated Reporting Council is a global coalition of regulators, investors, companies, standard setters, the accounting profession and NGOs, whose mission is to establish integrated reporting as the norm in public and private sectors (IIRC, 2018).

notion that the terminology is not well defined. In the executive summary, the authors attempt to define their terminology. Concluding with no distinction between terms such as ESG and SRI, they rationalize that it helps articulate their work in the broadest sense. The GSIA’s collective sustainability terms encompass:

- 1) Negative screening
- 2) “Best-in-class” screening
- 3) Norms-based screening
- 4) Integration of ESG factors
- 5) Sustainability themed investing
- 6) Impact/community investing, and
- 7) Corporate engagement and shareholder action.

Being able to get the message across as broadly as possible makes sense from GSIA’s standpoint. The alliance’s mission is to improve the visibility and impact of sustainable investing organizations globally. Excluding large shares of the market based on semantic technicalities seem counter-productive.

However, as the purpose of this study is to analyze a rating methodology that was constructed with ESG-factors in mind specifically, it feels expedient to distinguish between terms. For the purpose of this study, the umbrella term, encompassing both SRI and ESG among others will be “sustainable investing” or derivatives of the phrase. The interpretation of the collective term is kept consistent with GSIA’s seven categories named above.

Defining “SRI”

SRI is arguably the oldest and most common term within sustainable investing. It is also the term with the broadest variety of interpretations. I earlier introduced SRI as “socially responsible investing,” which it historically is known as, but there are other definitions with the same abbreviation. US SIF creates a broader definition by defining SRI as “sustainable, responsible and impact investing” (US SIF 2019).

As mentioned in section 3.1 and exemplified above, SRI is often used as a collective term for all sustainability-assets. When referring to the history of SRI however, many mention its origins as the 1700s and cutting financial ties with supporters of slavery practices. Modern era SRI gained momentum in the late 1960s (Richardson, 2008). In the following decades, the

appeal broadened through acts such as anti-violence campaigns, shareholder resolutions against napalm production and withdrawal of funds from banks invested in South Africa, just to name a few. Here, the common theme is negative screening.

In the academic literature and fund management, SRI is frequently applied in an exclusionary context as well. In a meta-study on sustainable investing, Fulten et. al. (2012) evaluate fund performance based on more than 100 academic studies. The findings show clear correlation between SRI and negative screens in the literature. Though the studies themselves might not have focused on negative screens, the funds in question were categorized as SRI and consistently weighted towards exclusion. “SRI-funds” are usually offered as a way for investors to keep their financial means invested in a consistent manner with their held beliefs and therefore exclude certain industries, not for financial performance, but to serve a demand.

In this thesis SRI will be defined as a negative screening process based on sustainability values and considerations. Among the categories defined as sustainable investing, SRI would cover “negative screening” and “norms-based screening” under this definition.

Defining “ESG”

While “ESG” is yet another case of overlapping term use in the literature, the wording “environment, social and governance” provides some guidance. The three components should be considered. More importantly, the rating methodology I use in this study rates companies compared to their industry peers and funds to their category peers, without exclusion. The rating is based on ESG analysis. It follows logically that I adhere to the methodology’s fundamentals. Therefore, in this paper, ESG will simply be defined as the incorporation of ESG-factors in financial analysis. Further, it implies positive screening, emphasizing environmental, social and governance factors. It does not exclude negative screening of industries in individual funds, but when present, negative screening is a bi-product rather than the intention. This definition finds support in literature and rating methodologies as well, and was seemingly the intention of the concept’s origins, or at least the result.

ESG funds

I briefly want to mention the term “ESG fund”. Over the past decade or so, the mutual fund market has increased its supply of portfolios in this category to tap into demand. The portfolio holdings of such funds have been (and still are) vastly diverse. Some funds apply different

screens, from negative to positive, while others simply create a portfolio of companies with high firsthand or third-party sustainability scores. Under a group of funds deemed “socially responsible” by Morningstar, several examples of this can be observed in ETFs and mutual funds. Vanguard **ESG** U.S Stock ETF seeks to track the FTSE US All cap Choice Index (Vanguard, 2019). Under their product summary, they state that as a part of their investment strategy, they “Specifically exclude stocks of companies in the following industries: adult entertainment, alcohol and tobacco, weapons [...]” While introducing the ETF as “ESG”, the fund predominantly aims to exclude a list of industries. JHancock **ESG** Large Cap Core Fund builds its strategy around investing in companies with “positive ESG practices” (John Hancock Investments, 2019). iShares **ESG** MSCI EM ETF which wishes to “Obtain exposure to higher rated environmental, social and governance (ESG) companies [...]” (iShares, 2019). These are all examples to show how the term “ESG fund” provides no concrete meaning.

By my definition, there are no “ESG funds”, as ESG investing is simply an approach that considers environmental, social and governance criterion. There is no absolute cut-off that deems a fund “ESG-fund” worthy, but rather a subjective preference.

3. The Morningstar Sustainability Rating. An overview

When deciding on which methodology to use for ESG ratings in the analysis, the two main options were MSCI ESG Ratings and the Morningstar Sustainability Rating. These are undoubtedly the two most comprehensive methods in the market for ESG ratings at portfolio level. The biggest difference between the two is that MSCI gathers its own company level data, while Morningstar receives it via Sustainalytics. In the end, the choice was more or less arbitrary as both companies seem to capture the essence of ESG performance despite slight differences in the information gathering process (Morningstar 2017). That being said, the methodology used to rate and rank funds in this paper is the Morningstar Sustainability Rating. In this chapter, I will briefly introduce the companies which are behind the development of MSR, followed by a detailed description of the rating methodology.

3.1 The Morningstar- and Sustainalytics collaboration

Morningstar is an independent investment research company that analyzes financial instruments for all levels of investors. The company's creator, Joe Mansueto founded Morningstar Inc. for all investors to have access to the same information as financial professionals (Morningstar, 2019). Today the company has over 5.000 employees operating in 27 countries. Through thirty-five years in the industry, Morningstar has made its presence known and is now one of the leading providers of investment research and management. In recent years Morningstar increased its value to investors by providing sustainability ratings through their partnership with Sustainalytics.

Sustainalytics is a global leader of ESG and Corporate Governance research and ratings (Sustainalytics, 2018). The company specializes in rating companies based on environmental, social and governance factors, helping investors facilitate their risk analysis. For instance, Sustainalytics raised governance concerns regarding Volkswagen AG several months before they admitted to cheating on emission tests in 2015 (Reuters, 2017).

In August 2015, Sustainalytics officially entered a strategic collaboration with Morningstar. Nearly two years later, in July 2017 Sustainalytics announced that Morningstar had acquired a 40 percent ownership stake in their company (Sustainalytics, 2018). The collaboration resulted in the development of Morningstar Sustainability Rating.

During the first quartile of 2016, Morningstar launched the Morningstar Sustainability Rating. Through the implementation, Morningstar established themselves as the premier provider of

sustainability ratings at fund level. Before MSR, investors with sustainable investment preferences had the option to buy ESG or SRI-specific mutual funds. Such funds are specifically created in order to cater to these investors demands. Competing in a shallow market (<2% of mutual funds), these funds have inherent limitations, such as increased costs, general distortionary effects and no reliable way of measuring actual sustainability. After the introduction of MSR however, over 20.000 funds have received sustainability ratings, opening a whole new world to sustainability-driven financial investors.

3.2 Rating methodology

The Morningstar Sustainability Rating rates funds from 1 (worst) to 5 (best) compared to their category peers. The rating is derived from a Portfolio Sustainability Score, consisting of the Portfolio ESG Score and the Portfolio Controversy Score. The portfolio level ESG- and controversy scores are derived from an aggregated average of scores in the underlying assets' company ESG- and company controversy scores.

The company ESG Score is a measure of the companies' ESG practices relative to their respective industries peers. Companies with sufficient available data are assigned a numerical, normalized ESG score. To derive the company ESG score, Sustainalytics provides Morningstar with company ratings that are peer group specific on a 0-100 scale (Morningstar, 2017). The ratings are based on more than 70 industry-unique indicators, divided into:

- Preparedness: Measuring management and policies to counter ESG related risks.
- Disclosure: Evaluating reporting practices and transparency regarding ESG issues.
- Performance: Both quantitative and qualitative measures of ESG performance.

Before publishing an ESG report on a company, Sustainalytics sends a draft to the company in question to gather feedback and add or update their information (Harvard, 2017).

An implication of Sustainalytics' approach is that a company's score can signal a percentile rank in its peer group which differs from a company with an identical score from a different peer group. Doyle (2018) exemplifies this when calculating the average ESG score for 4150 companies rated by Sustainalytics. He finds that, for example the utilities industry has a 61 ESG score average, while the healthcare industry average is 48. This juxtaposition illustrates a difference caused by industry specific indicators and show how scores don't necessarily

reflect sustainability. To make ESG scores comparable between groups, they are normalized through a two-step process. First, Morningstar applies a z-score transformation:

$$Z_c = \frac{ESG_{CS} - \mu_{PG}}{\sigma_{PG}}$$

Where

ESG_{CS} is the company ESG score provided by Sustainalytics,

μ_{PG} is the mean ESG score in the peer group, and

σ_{PG} is the standard deviation of ESG scores in the peer group.

Scores are then normalized with a theoretical range of 1-100. The average score is set to 50 for the respective industries. A company ESG score of 10 points above/below the average implies a score of 1 standard deviation above/below said average.

$$ESG_C = 50 + 10Z_c$$

Where

ESG_C is the Company ESG Score used to calculate Portfolio ESG Scores.

As such, the company Z-score describes how many standard deviations from the mean the respective company's ESG score is.

The company controversy score accounts for the company's most serious current controversy. If a company is involved in several controversies simultaneously, only the most prominent is included in the calculations. Sustainalytics' analysts measure the company's impact on environment and society as well as the risk to the company itself. The analysis is based on daily news monetization of 60.000 sources, covering more than 10.000 companies (Sustainalytics 2018). Examples of incidents which may result in an increased controversy score are oil spills, lawsuits, relation to controversial business partners or other incidents which negatively impact a company's reputation. In cases of severe controversies Sustainalytics contacts companies directly to refrain from large deviations between the controversy score and absolute impacts of a controversy. The MSR deduction resulting from

controversies ranges from 1 to 20, divided among six (five)⁸ categories on a hurricane scale,⁹ portrayed in the table below.

Table 2.1. Sustainalytics’ company controversy score

Controversy	Impact on environment or society	Risk to company	Controversy Score
5	Severe	Serious	20
4	High	Significant	16
3	Significant	Moderate	10
2	Moderate	Minimal	4
1	Low	Negligible	0.2
0	No evidence of controversy	None	0

Source: Morningstar, 2017

With normalized company ESG and controversy scores at hand, Morningstar applies the data to construct portfolio scores. The Portfolio Sustainability Score consists of the Portfolio ESG Score and the Portfolio Controversy Score. It can be defined as:

$$PSS = ESG_p - Contr_p$$

Where

PSS = Portfolio Sustainability Score,

ESG_p = Portfolio ESG score, and

Contr_p = Portfolio Controversy Score.

The Portfolio ESG Score is an asset-weighted average of all company ESG scores included in the portfolio:

$$ESG_p = \sum_{i=1}^n w_i ESG_{C_i}$$

⁸ Sustainalytics operates with 5 categories, as they exclude category “0”, while Morningstar lists 6 categories. This has no implications in practice.

⁹ Scale where severity (score deduction) increases exponentially with categories.






For a fund to receive a Portfolio ESG Score, Morningstar sets a minimum coverage threshold of 67%, increased from 50% in October 2018. This means that a fund does not get rated if more than a third of the underlying securities lack ESG scores. The funds ESG score is later rescaled to 100%, implying that there is no distinction between funds with 67% and 100% asset ESG coverage. While this could have some implications in theory, it is unlikely to be significant in practice.

In a similar fashion, the Portfolio Controversy Score is aggregated by the weight of the assets and their respective company controversy scores.

$$\text{Contr}_p = \sum_{i=1}^n w_i \text{Contr}_{p_i}$$

Finally, funds are arranged in their corresponding Morningstar Global Category¹⁰ and the Morningstar Sustainability Rating is derived. For a fund to receive an MSR, it is required that its corresponding category consists of at least 30 funds. Based on the funds’ numerical score, they are awarded “globe”-ratings from 1-5, with an approximate normal distribution internally in the category. The distribution is illustrated in table 3.2.

Table 3.2 Morningstar Sustainability Rating distribution

Distribution	Description	MSR rating icon
Highest 10%	High	
Next 22,5%	Above average	
Middle 35%	Average	
Next 22,5%	Below average	
Lowest 10%	Low	

Source: Morningstar, 2017

¹⁰ The Morningstar Global Category classifications can be found here: <https://www.morningstar.com/content/dam/marketing/shared/research/methodology/860250-GlobalCategoryClassifications.pdf>

While the Portfolio Sustainability Score explains how the underlying companies fare in their respective industries, the globe rating measures how a fund compares to its category peers. As such, the two metrics have different functions. The score-rating functions as a cross-category rating system, used as an absolute measure of portfolio sustainability, while the globe-ratings are best used for comparing funds within a specific investment category.

To summarize, Sustainalytics provides Morningstar with data on company level regarding sustainability and controversies. Morningstar then applies the data to construct normalized values on company level that is used in creation of portfolio scores and the Morningstar Sustainability Rating.

3.3 Historical Portfolio Sustainability Score

Since launching MSR, Morningstar have continuously improved their rating methodology. One of the more recent changes is the incorporation of a historical sustainability score. As of October 2018, MSRs are based on Historical Portfolio Sustainability Scores (HPSS), rather than the most recent portfolio ESG and controversy values (Morningstar, 2018a). The historical score includes a weighted average of the portfolios’ holdings over the trailing 12 months, with stronger emphasis on more recent holdings. It is calculated using the following formula:

$$\text{Historical Portfolio Sustainability Score} = \frac{\sum_{i=0}^{11} (12 - i) * PSS_i}{\sum_{i=0}^{11} i + 1}$$

3.4 Environmental (E), Social (S) and Governance (G) scores

Morningstar also provides individual scores for each component of “ESG”, an environmental, a social and a governance score. Morningstar refers to these as “Portfolio Pillar Scores” (Morningstar, 2018b). The pillar scores are first disaggregated from Sustainalytics’ combined ESG score, then the same methodology is applied to normalize individual pillars within their peer group as with the ESG score described in section 3.2. To calculate a Portfolio Pillar Score, Morningstar uses the normalized company pillar scores, the weight of the assets in the portfolio as well as well as a “Peer Pillar Weight” provided by Sustainalytics. The Peer Pillar Weight is employed to account for differences in the contribution of each pillar in the overall

ESG score between peer groups. The individual pillar scores can be used to evaluate the importance of each factor in the ESG assessment.

4. Literature review

The literature regarding sustainability is extensive and has developed through the years. Keeping the terminology consistent, most studies are looking at performance of “SRI funds” rather than defining the investment approach. This leads to a broad variety of studies claiming to study the same topic. In recent years, the approach within “SRI” has been moving gradually from negative screens to positive screens (Fulten et al, 2012; Eurosif, 2018). It has also become a less confined term, as more investment styles are included. Studies referenced here do not necessarily account for this shift.

4.1 Sustainability and performance

Nofsinger and Varma (2014) study the performance of U.S.-based mutual funds with ESG-driven attributes during periods of financial distress. The study concludes that these funds outperform conventional funds. However, this comes at the cost of performance during non-crisis periods. The authors attribute this pattern to “ESG funds that use positive screening techniques.” Motivated by Nofsinger and Varma (2014), Lesser et al (2016) expand the geographical scope to include international funds. Their study reveals no outperformance regardless of market crises and conclude that results from Nofsinger and Varma (2014) are not transferable to international markets. The authors suggest that the outperformance found by Nofsinger and Varma are due to superior active management by U.S. fund managers during market turmoil.

In a study on ESG factors’ relation to risk-adjusted stock performance, Kumar et al (2016) compares 157 companies listed on the Dow Jones Sustainability Index (DJSI) with 809 that are not. The authors create industry-specific portfolios based on the stocks in DJSI in order to eliminate biases that come from industry ESG performance rather than company performance. The reference group was sorted by industries accordingly. With this approach, and weekly returns over a two-year period (2014-2015), they find that for the 12 industries in question, the ESG group had better returns in 8. Across all 12, the ESG group showed 6,12% higher annualized equity returns on average. In terms of risk-adjusted returns, the ESG group

outscored the reference group in 9 out of 12 industries both in terms of Sharpe¹¹ and Treynor¹² ratio.

Improved risk-adjusted returns were also found by Eccles et al (2016). In their study, the data is divided into two main investment universes, a “Global All” and a “Global Developed Markets.” The former consists of roughly 85% of all investable equities globally, while the latter includes 85% of all developed market equities. From the two universes, six portfolios were created:

- 1) Global All unscreened
- 2) Global All top 10% ESG
- 3) Global All top 25% ESG
- 4) Global Developed Markets unscreened
- 5) Global Developed Markets top 10% ESG
- 6) Global Developed Markets top 25% ESG

When compared to their unscreened counterparts, the authors find that for three out of four portfolios, ESG screening contributes to risk-adjusted excess returns. Portfolio 6) created a small negative alpha. For the other three ESG screened portfolios there was an increase in risk that was offset by sufficient returns to create a positive alpha. Additionally, analyzing daily return distributions on stock-level, the authors find that ESG screening reduces tail-risk. The analysis was done by measuring number of observations below -3 standard deviations.

Van Beurden and Gössling (2008) conclude a literature review saying that “Good Ethics is Good Business.” The study reviews the relationship between Corporate Social Reasonability (CSP)¹³ and Corporate Financial Performance (CFP) and found a generally positive relationship. 68% of studies in the analysis showed a significant, positive relationship, while 6% were negative. The rest were non-significant. Throughout the reviewed studies, there was a persistence of the size-factor as a confounding variable. The authors acknowledge the lack of an industry wide definition of CSP, making it harder to draw definite conclusions.

¹¹ The Sharpe ratio is commonly used measure of risk-adjusted returns. It is calculated by subtracting a risk-free return from the portfolio and dividing by the portfolio’s standard deviation.

¹² The Treynor ratio is another measure of risk-adjusted returns. It is calculated in a similar way to the Sharpe ratio, except the divisor is the portfolio beta.

¹³ According to Fulten et al (2012), CSP in the literature is essentially the same as ESG. Furthermore, regjeringen.no (2016) defines CSR as “business expected to assume responsibility for their impact on people, environment, and the communities and societies in which they operate”, supporting the relation to ESG.

Adler and Kritzman (2008) hypothesize that there must exist a negative relationship between sustainable investing and performance. Through Monte Carlo simulation, the authors find that imposing sustainability constraints on the portfolio carry additional costs (0,17% to 2,4%) to the investment approach, in accordance with the Efficient Market Hypothesis (EMH) (Fama, 1970). Similarly, Renneboog et al (2008) find that performance of mutual funds suffers when enforcing ethical considerations. Though generally not statistically significant, the authors display benchmark-relative underperformance of 2,2% to 6,5% for a global basket of mutual funds. Additionally, they find statistically significant, negative tilting towards the market factor in their difference portfolio, implying lower sensitivity to market risk when compared to conventional funds.

In an update from their previous study, Johnsen and Gjørlberg (2008) find that SRI-funds have delivered substantially lower risk-adjusted returns than conventional benchmarks based on data from 2003 to 2007. For this sub-period in the study, the SRI-indices carried a negative, significant alpha between -0,78 and -2,60 at minimum 10% significance level. The authors partially attribute the skewness in benchmark-relative underperformance to a positive selection process, imposing constraints on the investment universe. Furthermore, they find a systematic prevalence of factor-tilting towards size¹⁴ where SRI-funds contain disproportionate amounts of large capitalization companies.

Bauer et al (2005) study the performance of an international database of ethical mutual funds compared to conventional funds. After controlling for investment style, the authors find no evidence of differences in risk-adjusted returns. The ethical portfolios were constructed of mutual funds investing in assets using ethical screens in each country included in the study. The reference “conventional funds” group was gathered using all other equity mutual funds in the respective countries. In their CAPM-based single index model, no alphas were found significant when compared to the conventional funds group. All market betas were found to be <1 and generally lower in the ethical funds group, implying lower sensitivity to market risk. When adjusting for factors such as size, value and momentum with the Carhart 4-factor model, alphas maintained their pattern of being statistically insignificant. Consistent with the single-factor analysis, ethical mutual funds generally showed lower exposure to market risk in the multi-factor model. For other systematic risk factors, there were several statistically significant coefficients, but with changing signs and seemingly no exhibited patterns. When

¹⁴ Using Fama and French 3-factor model.

dividing into sub-periods, the authors found that ethical mutual funds provided superior risk-adjusted returns between 1998 and 2001, after performing worse at the start of the 90s. They attribute this to a “catching-up phase.” Johnsen and Gjølborg (2008) instead attribute the period of excess growth to loading on the size-factor and the “dot.com-bubble.”¹⁵ As mentioned, SRI funds were found to contain a large cap bias, and large companies received especially high pricing during the dot.com-bubble period of 1997-2000.

4.2 Meta studies on sustainability and performance

In a comprehensive meta-analysis, Friede et al (2015) find a nonnegative relationship between ESG criteria and CFP in 89% of reviewed papers. The analysis is based on roughly 2200 empirical studies. 1816 studies were found to analyze the relationship between ESG and CFP, of which 48,2% showed a positive market-based outperformance, 10,7% were negative while 23,0% and 18,0% of studies were neutral or mixed respectively. The study concludes that the ESG-CFP performance relationship is mostly positive on company level and neutral or mixed at portfolio level.

Through analyzing more than 100 academic studies, Fulten et al. (2012) show that funds categorized as “SRI” tend to rely on exclusionary screens. The performance based on this approach was overall regarded as neutral, though there was a mixture of results. When disaggregating and screening for ESG ratings, the authors found that 89% of studies exhibit market-based outperformance. The most important factor for performance when screening for ESG was found to be the governance-factor, followed by the environment and social factors. This is consistent with later findings by Han et al (2016) when studying the relationship between ESG and performance in Korean markets. Regarding risk, 100% of studies in question showed a lower cost of capital for companies with high ratings of ESG, implying that the increased performance comes from lower risk taking.

In an enhanced meta-study, Clark et al. (2015) analyze results from more than 200 academic studies and sources. Similar to results found by Fulten et al (2012), the study suggests that companies scoring well in terms of sustainability show better performance and less associated

¹⁵ A market surge at the turn of the millennium. The surge come to as a function of the market’s overenthusiasm for tech companies and IPOs. The bubble burst at the start of the millennium, leading to a steep decline in the stock market.

risk. Based on 41 studies on performance of ESG or its components, 80% showed market-based outperformance. Results are portrayed in table 4.1.

Table 4.1. Results from Clark et al (2015) of ESG studies and market outperformance.

	n	Positive	Negative	Mixed	Neutral
ESG	12	10	1	1	0
ES	2	1	0	1	0
E	8	7	0	1	0
S	8	6	0	2	0
G	11	9	0	0	2
SUM	41	33	1	5	2

Source: Clark et al (2015)

4.3 Morningstar Sustainability Rating and mutual fund performance

In a recent study, forthcoming in the Journal of Accounting and Finance, Chatterjee et al (2018) study the effects of the Morningstar Portfolio ESG Score on 73 U.S. mutual funds. The authors divide the funds into three categories, High ESG (top 33%), Mid ESG (next 33%) and Low ESG (bottom 33%). Through cross-sectional analysis, controlling for variables such as ESG ratings, tenure, expense ratios and the funds' age, they find that generally, lower and mid rated funds perform better than high rated funds, both in terms of absolute returns and risk-adjusted returns. However, in accordance with previous literature they find better risk-adjusted returns in higher rated funds in periods of market downturns (2005-2008). They also find evidence that high sustainability scoring funds mimic the performance of funds with an SRI-mandate. The study uses a cross-sectional variable in Morningstar's ESG ratings and applies it to the 12-year period 2005-2016. This could have some implications on the results if the variable is not persistent for the period.

Similarly, Dolvin et al (2017) study the relationship between a Morningstar sustainability metric and U.S. domiciled mutual funds. A difference from the study by Chatterjee et al (2018) is the incorporation of Morningstar Portfolio Sustainability Score, rather than the ESG score in the performance analysis. The average Portfolio Sustainability Score in the full data sample was 43,98 with individual fund scores ranging from 37,22 to 53,79. The authors report a significant difference in sustainability scores between small-cap and large-cap funds in their data, where small-cap funds have lower scores. This implies a large-cap bias when employing the sustainability score as only screening criterion for an investors fund selection. In the study, the authors segment their sample into quintiles to measure performance of top 20%,

middle 60% and bottom 20%. They find that in the bottom quintile, 95% of funds are small-cap funds while in the upper quintile, 93% are large-cap funds. Thus, they predict (and find, significant at 1% level) a prevalence of size loading in the multi factor regression analysis. Further in the study, the authors investigate how funds with a self-proclaimed SRI mandate score in terms of Morningstar's metrics. Generally, these funds are in the top quintile, though 35% are outside the top 20% of sustainability scores. However, with $n=31$, the sample is too shallow to draw robust conclusions.

In a master's thesis from the Norwegian School of Economics, Auran and Kristiansen (2016) study the relationship between Morningstar Sustainability Rating and risk-adjusted performance in mutual funds with at least 50% of portfolio value invested in European developed markets. The authors define their high and low portfolios as top 10% (5 globes) and bottom 10% (1 globe) MRS respectively. Through multifactor regression with the Fama and French 3-factor model and the Carhart 4-factor model, they find no evidence of a risk-adjusted return difference between high and low rated funds. The results show that high sustainability funds generally load more on the market factor than low sustainability funds, i.e. statistically significant higher beta. Furthermore, the authors find a negative, statistically significant difference in loading on SMB, implying that high sustainability funds load more on large capitalization stocks than their low sustainability counterparts. It is worth noting that the study employs the globe ratings to rank funds. In chapter 6 of this paper, I argue why I think the globe-rating is an inaccurate measure of sustainability in this context.

4.4 Controversies, controversy scores and fund evaluation

Nofsinger et al (2016) study the relationship between institutional investors' preferences and controversial products. The authors find that after the Family Smoking Prevention and Tobacco Control Act¹⁶ of 2009, institutions dramatically reduce their tobacco-stock holdings and end up underweighting tobacco firms. Similarly, firearms firms were underweighted by institutions until the Federal Assault Weapons Ban¹⁷ expired in 2004. The study also finds that institutions underweight products that carry controversy in general, like alcohol and

¹⁶ The U.S. Food and Drug Administration (FDA) were in 2009 given the authority to regulate the manufacture, distribution and marketing of tobacco products in the "Tobacco Control Act" (FDA, 2018)

¹⁷ A 10-year prohibition of the manufacture of 18 specific firearms for civilian use by the congress from 1994 (Washingtonpost, 2012).

nuclear stocks. They conclude that these discriminations are not based on social issues, but rather economic incentives.

Utz (2017) makes a reference to Chatterjee et al (2009) when he sums up reasons for why an investor might opt for a “socially responsible way” to allocate his or her money. The conclusions based on the list is that motives generally lie in reducing exposure to corporate scandals - or reducing tail risk. In the study, Utz analyzes the reliability of ESG assessments in predicting corporate controversies and scandals. The author finds that controversy scores in general are bad indicators of future scandals, but they are a well-functioning metric in the aspect of retrospective indication.

Nofsinger and Varma (2014) point to several cases of controversies that have led to long-term underperformance. An example is BP’s Deepwater Horizon oil spill from 2010. After the initial oil spill, the stock price fell victim to a two-month downswing, totaling a 55% decline. During the subsequent five years, the stock went on to underperform the S&P 500 as a function of high settlement costs. The authors argue that firms with high probability of “adverse ES events” (controversies) may have lower long-term returns because of an underestimation regarding possible future losses.

4.5 Critique of ESG ratings as objective measures

With a lack of standardization in reporting and measuring, ESG ratings are unquestionably a flawed metric. As mentioned above, in 4.3, a study on the U.S. mutual fund market found that 93% of the upper quintile sustainability scores were large cap funds. A large cap bias has been a recurring theme while studying the literature (van Beurden and Gössling (2008); Johnsen and Gjørlberg (2008); Dolvin et al (2017); Auran and Kristiansen (2016)).

In an article posted by the American Council for Capital Formation (ACCF), the author Timothy Doyle highlights some issues regarding the policies of current ESG rating agencies (Doyle, 2018). Summarizing his study, he makes claims of several categories of issues and biases leading to disparities in individual ratings:

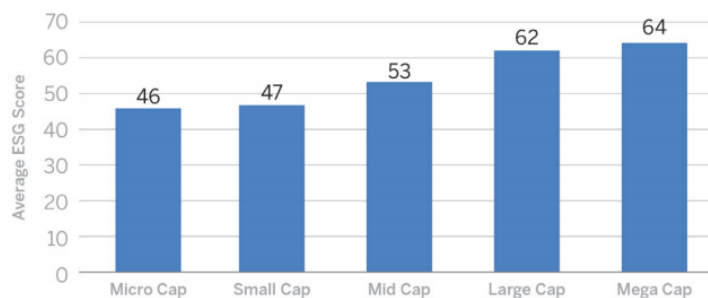
- Disclosure limitations and lack of standardization
- Company size bias
- Geographic bias
- Industry sector bias

- Inconsistencies between rating agencies
- Failure to identify risks

Doyle claims that in general, ESG rating systems reward companies with greater disclosure. This leads to the possibility of a company with weak ESG practices (but robust disclosure practices) outscoring a peer with stronger ESG practices (but weaker disclosure practices). He argues that this policy creates a situation where the emphasis is on disclosure, rather than the underlying, disclosed risks. In the article, this is exemplified with the Goodyear Tire & Rubber company, which by Sustainalytics is scored 15 points ahead of the industry average. The score is in spite of ESG issues, fines and settlements. The conclusion is that the comprehensive disclosure practice is to “blame” for the high scoring.

Analyzing the scores of over 4000 Sustainalytics rated companies, Doyle presents the following table:

Table 4.2: Average ESG Score by market cap



Source: Doyle (2018)

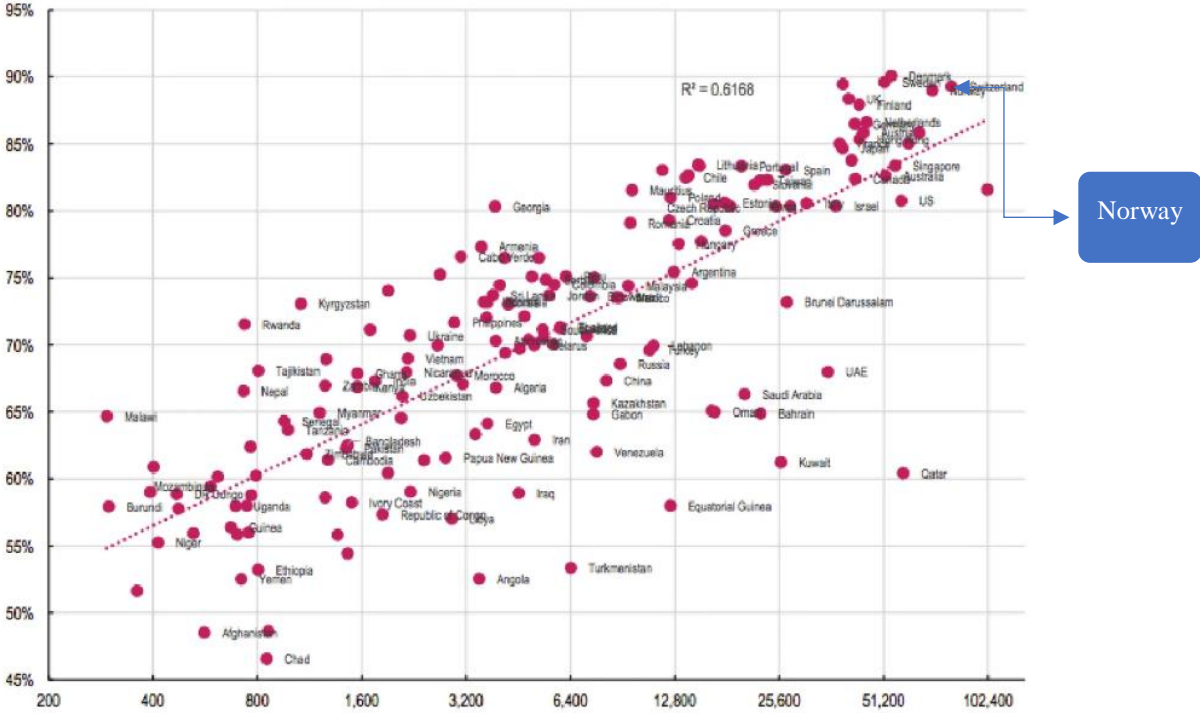
Table 4.2 shows how sustainability ratings scale with capitalization for 4150 Sustainalytics rated companies. The author does not specify the period for which the data in question covers.

Seemingly, small and mid-cap companies are at a disadvantage when agencies assign ESG scores. The large-cap bias somewhat synergizes with the point regarding disclosure, as larger companies are better prepared to handle the costs associated with full transparency.

The juxtaposition of the two largest ESG continents (Europe and North America) with their respected ESG scores highlights a geographical bias in the distribution of sustainability ratings. With a reference to the same data as above, the average Sustainalytics score for companies in Europe is reported to be 66, while the American peers’ score 50. The increased scores in Europe in general can be a product of stricter disclosure requirements. Reinforcing the notion of a geographical bias, Dolvin et al (2017) analyses a sample of 1853 U.S. domiciled mutual funds, scoring an average of 43,98 on the Morningstar Portfolio

Sustainability Score. In comparison, my data covering 215 Norwegian domiciled funds average 51,11 on the same metric. This very notable difference can be explained further by an article posted in Financial Times. The article references Renaissance Capital, portraying the following overview of ESG scores and geographical rankings:

Figure 4.1: Renaissance Capital ESG score vs GDP per capita



Source: Financial Times (2018)
 Figure 4.1 shows how ESG scores tend to scale with GDP per capita.

From figure 4.1 it is apparent how western countries are more likely to receive high ESG scores. ESG scores scaling with GDP per capita is partially covered by Doyle (2018) as he attributes certain biases to the disclosure practice. While Europe is outperforming North America on a continent level, Norway is one of the top performing nations globally.

The industry sector bias refers to how certain industries have higher average ratings than others. This is often a function of industry specific indicators, as those mentioned in section 3.1. Morningstar attempts to solve this by normalizing scores. Regarding the different agencies, they were found to have significant inconsistencies in company ratings. Data referenced by Doyle (2018) tell tales of a 0,32 correlation between MSCI and Sustainalytics' ratings for the companies in S&P 1200. However, without further information, the correlation provides little to the discussion in my opinion. As the rating agencies score companies relative

to their industry and the S&P index contains a variety of industries, the correlation is only interesting if it is presented as an aggregate of all individual industries' correlations.

Lastly, Doyle (2018) emphasizes the failure of ESG scores in accurately identifying risks. To reinforce this claim, he points to corporate scandals and how many affected companies have had above average scores before the scandal took place. This is consistent with findings by Utz (2017) as he concludes that ESG measures are unreliable in predicting scandals.

5. Theoretical framework

5.1 Sustainability and asset pricing

Modern Portfolio Theory (MPT) (Markowitz, 1952) suggests that imposing constraints on the investment portfolio results in an efficiency loss, assuming a risk averse investor. Any screening process comes with inherent limitations of the investment universe, i.e. constraining the portfolio. Assuming the MPT holds true, why would an investor then opt to screen for sustainability in his portfolio?

Hill et al (2007) proposes that some groups of investors simply know what they want to exclude from their investment strategy. These groups are typically bound by their inherent values, e.g. religious or political beliefs. Assuming efficient markets, purely moral values are likely to jeopardize some performance. However, for these groups, the added utility that stems from investing consistently with their held beliefs outweighs any performance loss resulting from this approach.

While some investors want a portfolio based on specific ethical or moral values, other investors just want high ESG profile (Hill et al, 2007). Contrary to the ethical approach, ESG-focused investing comes with the assumption that sustainability can have financial relevance.

When the United Nations Global Compact published “Who Cares Wins,” it was stated that a “growing body of empirical evidence” indicated that ESG factors are contributory to “better risk-adjusted financial performance than their industry peers” (Knoepfel, 2004). Resulting recommendations say that investors should improve their understanding of the link between ESG performance and value creation and consider ESG issues a potential source of competitive advantage.

Studying performance of mutual funds, the main objective is to determine what is generating “abnormal” performance, where abnormal is defined as in excess of the market and other systematic risk factors. Fama (1970) claims that this is either done by (a) having monopolistic access to special information, or (b) being better at uncovering such information, i.e. better understanding of the implications of publicly available information. In both cases there is a level of market efficiency implied, strong or weak form respectively. In the referenced paper, Fama presents the notorious Efficient Market Hypothesis (EMH). If an investor believes in

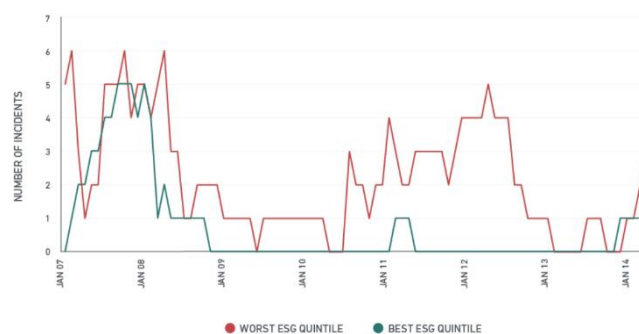
the presence of market efficiency as presented in its semi-strong form¹⁸, there would be no abnormal performance gained from screening one's investment options. The EMH plays a fundamental role in financial theory, but it is not without its critics. Perhaps the most famous critic is Robert J. Shiller, whom (accompanied by Eugene Fama) in 2013 received the Nobel prize for his empirical work on market (in)efficiencies. Shiller (2002) refutes the EMH, stating that Fama's work incorrectly reflects psychological underpinnings of behavioral finance. First and foremost, Shiller points to serial dependencies in stock market returns, especially related to large swings. Fama explains these as a part of efficient markets, as they convey the markets *current* interpretation of information. While a significant movement in asset price is usually followed by another, the subsequent movements do not have to follow a pattern, i.e. successive large movements often switch signs. Fama uses this to reinforce the notion of the EMH, while Shiller says that this is exactly what one would expect in an inefficient market.

Screening for ESG is typically done under the assumption that sustainable assets are less exposed to excessive movements in prices. More specified, a company exceeding in environmental, social and governance aspects is less likely to experience downside risk compared to worse ESG-scoring peers. If true, the EMH would state that this would come with a sufficient reduction in expected returns. So far in this paper, there are several illustrated examples of this being both true and not true.

Nofsinger and Varma (2014) suggests that the market underestimates future downside risk. Most historical data shows that the investment market is influenced by some positive excess kurtosis and negative skewness, generally known as high peaks and fat tails, with a longer left-side (downside) tail. MSCI reports a lower frequency of idiosyncratic risk incidents in high rated companies (MSCI, 2018). Conversely, low scoring companies experience more major drawdowns.

¹⁸ In its semi-strong form, the EMH suggests that all publicly available information is reflected in current asset prices.

Figure 5.1. MSCI top vs. bottom quintile idiosyncratic risk incidents 2007-2014.



Source: MSCI (2018)

From the figure above, it is apparent how the low-end ESG quintile is more exposed to idiosyncratic risk incidents (fraud, corruption, litigation etc.). An idiosyncratic risk incident is in this case defined as drawdown of >95% or bankruptcy. The incident-corresponding ESG-rating is defined as the rating before the drawdown period begun. This can be explained by an existing positive relationship between ESG characteristics and risk control (Giese et al, 2017). With better standards for risk control, high ESG-rated companies experience a lower frequency of serious firm-specific incidents. A consistent pattern of high ESG scores and low idiosyncratic risk is emerging through academic work (f.ex. Bauer et al 2009; Oikonomou et al, 2011; Lee and Faff, 2009) and seems to extend to most markets. Furthermore, higher rated companies are found to carry lower cost of capital (Bauer et al, 2009; Fulten et al, 2012; Oikonomou et al, 2011), which can factor into long-term company-specific sustainability. Ultimately ESG considerations are found to act like an insurance from idiosyncratic downside risk.

Economic theory tells us that an investor is not compensated for taking on increased idiosyncratic (or diversifiable) risk. Can ESG create increased risk-adjusted returns by taking advantage of the markets underestimation of idiosyncratic risk connected to ESG-factors?

5.2 Performance measures and systematic risk factors

Measuring financial performance, one is interested in more than the expected return of the investment. Investing money in mutual funds comes with inherent risk. In a theoretical world with risk-neutral investors or infinite money, the level of exposure to systematic risk would not play a role in performance measures. However, according to behavioral finance, the average investor is risk averse (and does not possess infinite capital). Risk aversion insinuates that if returns are equal, investors will choose the option with lower associated risk. In some

cases (typically more extreme cases), investors are also willing to sacrifice expected returns to align with their utility regarding loss-aversion. Thus, performance is measured by analyzing the relationship between expected return and associated risk.

But what risk? As mentioned above, an investor is not compensated for taking on idiosyncratic risk, also known as unsystematic, firm-specific or diversifiable risk. The latter synonym, “diversifiable risk” indicates that such risk can be removed through sufficient diversification. Imagine two investors, one diversified, the other not. When investing in the same asset, the first investor does only have to worry about the systematic risk factors, as when this particular asset performs poorly, he is compensated in other aspects of his portfolio. The second investor is subjected to a larger portion of total risk and therefore should require a lower price. However, as there are enough diversified investors in the market, an asset is priced in accordance with its associated systematic risk factors. Systematic risk factors can be avoided by an individual, but someone must always undertake them and is therefore compensated.

The capital asset pricing model (CAPM) (Sharpe, 1964; Lintner, 1965) is the cornerstone of empirical analysis. The model was built on the seminal work of Markowitz (1952) on diversification, and describes the relationship between an investment’s expected return in relation to its associated market risk:

$$E(r_i) - r_f = \beta_i * (E(r_m) - r_f)$$

Where

$E(r_p)$ is the expected return of the investment,

r_f is the risk-free investment option,

β_i is the sensitivity of investment returns with respect to market returns, calculated as the covariance between the investment and the market divided by the market variance, and

$E(r_m)$ is the expected market return.

Fama and French (1993) expand on CAPM, as they find that univariate market-based models do not adequately embody common risk factors in the stock market. In their paper, Fama and French introduce two new explanatory variables of expected stock returns, namely the SMB and the HML-factors. The first factor, SMB, or *small minus big* relates to company size. The

factor encompasses the historical excess (average) returns of small-cap companies over large-cap companies. The factor is purely based on the relationship between returns and capitalization and is included in the model on the rationale that small-cap stocks tend to provide higher returns than large-cap counterparts. HML, or *high minus low* refers to a company's book-to-market ratio. The rationale behind this factor is that high book-to-market companies, or "value stocks" on average give higher returns than low book-to-market equities, or "growth stocks".

Combining the two above-mentioned risk factors with the market factor, this model is known as the Fama-French Three-Factor Model, formulated as:

$$E(r_i) - r_f = \beta_1(E(r_m) - r_f) + \beta_2SMB + \beta_3HML$$

In the multi-factor model, the β_i are factor coefficients that explain how the asset or portfolio is affected by market risk (β_1), capitalization (β_2) and book-to-market ratio (β_3). Recall from the literature review that sustainable funds have historically invested in large-cap stocks (Van Beurden and Gössling (2008); Johnsen and Gjølborg, (2008) Dolvin et al (2017); Auran and Kristiansen (2016)).

With the theoretical model established, we can employ historical data to estimate variables and create the following, practicable model:

$$r_i - r_f = \alpha_i + \beta_1(r_m - r_f) + \beta_2SMB + \beta_3HML + \varepsilon_i$$

Where

r_i is the estimated expected investment returns,

r_f is a measure for risk-free rate,

r_m is the estimated expected market return, and

ε_i is the error term, or residuals from the regression.

The α , often referred to as "Jensen's Alpha" (Jensen, 1967) is a measure of abnormal performance, i.e. performance not explained by the factor model. In this paper, I use Jensen's three-factor alpha as well as the Sharpe ratio (Sharpe, 1966) to measure risk-adjusted performance. The Sharpe ratio is defined as the expected return in excess of the risk-free rate relative to acquired risk. In practice, it is the earned excess return divided by the historical standard deviation. The multivariate regression is performed in Microsoft Excel, using

ordinary least squares (OLS).

5.3 Cross-sectional analysis

The models discussed and presented above, are regression models used in conjunction with time-series data sets. For the purpose of this study, time-series analyses are relevant as long as persistence is kept within the data. The nature of MSR as a variable is cross-sectional, referring to the fact that the availability of scores is limited to one point in time. This does not hinder the use of time-series analyses, but they must be handled with caution - a subject discussed further in section 6.4.

Given the vast variety of metrics and variables in this study, there are benefits found in analyzing the explanatory value of each metric on dependent variables such as fund returns, e.g. “How do fund returns change when sustainability ratings change?” To answer this, I supplement the study with a set of cross-sectional regression analyses, using average fund returns as the dependent variable, and possible descriptive characteristics as independent variables. Similar to the time-series analyses, this is done through OLS in Excel, using the following practicable model:

$$\bar{Y}_i = \alpha_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_j X_{ji} + \varepsilon_i$$

Where,

$X_{1i} \dots X_{ji}$ are independent variables explaining the changes in average returns.

6. Data

6.1 Data collection - describing the raw data

The data set used in this study consists of monthly simple returns for 146 Norwegian mutual funds, as categorized by Morningstar. The funds are categorized as Norwegian based on their International Securities Identification Number (ISIN) and the funds are domiciled in Norway. By Morningstar's definition, "Norwegian" does not limit the geographical investment area. The sample period is set to five years, stretching from January 2014 to December 2018. The determined sample period is a function of wanting a period long enough to draw robust conclusions, but not so long that lack of persistence in the data compromises results.¹⁹ The monthly return series was kindly provided by Client Service Consultant Christopher Greiner at Morningstar. Management fees have been extracted from returns. Fund returns or price data is generally available through a variety of open sources, such as Netfonds.

Sustainability ratings were also provided through the generosity of Morningstar. The sustainability scores and ratings are dated to November 30th, 2018. At Morningstar's webpage, most of the funds in this study have available Morningstar Sustainability Rating, (Historical) Portfolio Sustainability Score, Portfolio ESG Score and Portfolio Controversy Score. Other data, such as pillar scores are accessible through Morningstar's subscription based "Morningstar Direct."

The benchmark used in this study is the OSEFX. When deciding on a benchmark, the first decision was whether to opt for a global or Norwegian fund index. Despite the unlimited geographical investment universe, there is a heavy weight towards the Norwegian stock market in the analyzed funds. Also, this paper is first and foremost directed towards Norwegian investors. It therefore felt expedient to choose between Norwegian indices. Once limited to Norwegian indices, the most natural option (and decision) was the Oslo Børs Mutual Fund Index (OSEFX). Though the funds in this study are benchmarked against a variety of indices, OSEFX has a high recurrence as reference index in the data.

The OSEFX is a capped version of the Oslo Stock Exchange Benchmark Index (OSEBX). The weight-adjustments are in accordance with the Undertakings for Collective Investments In Transferable Securities (UCITS) for mutual funds (Oslobors, 2019). These include a

¹⁹ Persistence relates to keeping the results representative. See section 6.4 for more information.

maximum weight of a single security limited to 10% of total index market value. Moreover, there is a limit on the sum of all securities exceeding 5% of market value to 40% of said value. Daily net asset values (NAV) for OSEFX were extracted from Netfonds.

As mentioned above, fund returns are net returns after a deduction of management costs. While index investing is not free (index funds usually come with a yearly cost of roughly 0,3%), costs related to index-tracking are rarely accounted for in academic research. In compliance with previous research, I have not subtracted costs from benchmark returns.

The measure for risk free rate is a three-year Norwegian sovereign bond, issued by Norges Bank, extracted from Norges Bank's webpage. All mutual fund and benchmark returns are noted in excess of the risk-free rate.

The independent variables used in the multi-factor regression model, SMB and HML were downloaded from Kenneth French' website²⁰. The website offers a variety of investment market factors. Employing the American market factors is standard practice, though potentially problematic. As of writing this (spring, 2019) the American small-cap index, Russel 2000 has exercised a cumulative 5-year return of 35%. During the same timeframe the Oslo Børs Small-Cap Index has suffered a 4% downturn. My dataset consists of funds with a broad investment universe (though weighted towards Norway and other Nordic countries) and I therefore considered creating factors by category weights. However, due to the numerous amounts of portfolios with varying category weights, this would be very time consuming, and unlikely to yield effectual differences. After working with the data and sorting funds by capitalization categories, I found that the American market factors closer represent differences in returns in my data than the European developed market factors. I eventually opted to follow the standard practice in this regard and used the American factors for the analysis. I readily admit the potential problems with this approach but remind the reader that this is in fact the standard approach.

6.2 Filters on the dataset and fund categories

The unfiltered data set used in this study consisted of 215 mutual funds domiciled in Norway as described above. To avoid results being influenced by the funds' allocation strategy, referring mainly to percentage stock distribution in the funds, the data was filtered to be more homogeneous in that regard. First and foremost, all funds categorized as "fixed income" were

²⁰ Fama-French factors available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

removed. Secondly, any fund categorized by its allocation and not in the “aggressive” category were removed. Funds in these categories were allowed a stock allocation downwards of 50% (in “moderate allocation”) or lower (in “cautious allocation”). Lastly, all funds categorized by their “target date” were excluded from the sample as they generally adjust allocation strategy significantly over time. After filtering by category, there were 183 remaining funds. For reasons discussed in section 6.1 and further in 6.4, the sample period for the analysis is set to five years, from January 2014 to December 2018. Excluding funds that do not contain these 60 months of observations, I was left with 146 funds. While Morningstar doesn’t provide specific allocation requirements for all their categories, this should be a representative sample of funds that are predominantly invested in stocks. Summarized in the table below is the category distribution of the 146 funds in the sample.

Table 6.1: Overview of fund allocation by category.

Category	Europe Eq. Mid/Small	Global Eq. Large Cap	Europe Eq. Large Cap	Aggressive Allocation	Sector Equity	Global EM Equity	Equity Misc.	Other Equity	Sum
n	42	37	25	11	10	7	7	7	146

In the table, “Sector Equity” encompasses all funds invested in a single sector, e.g. healthcare (3 funds) or technology (1 fund). The “Other Equity” tab contains all other equity-labeled categories, represented by three or fewer funds individually which didn’t fit into any of the other categories. Examples are “Global Equity Mid/Small cap” (1 fund) and US Equity Large Cap Blend (3 funds). The remaining categories are official Morningstar Global Categories, for which classifications can be found in footnotes on page 19.

In terms of outliers, I have not trimmed neither fund return series, nor fund sustainability scores based on distance between observations. All sustainability metrics, with the exception of controversy scores have gone through a normalization process, making possible outliers small and not interfering with regression models. For the return series, deleting outliers from the dataset is expedient, if not necessary when working with stock returns. In a case of fund returns, portfolios are at least somewhat diversified, and outliers are in my opinion more valuable than damaging to the analysis.

6.3 Sustainability metrics in the analysis

The main purpose of this study is to look for a relationship between the Morningstar Sustainability Rating and risk-adjusted returns in mutual funds. Within MSR there are several

metrics which seek to capture sustainability, as portrayed in section 3. In this study, I look at a variety of these metrics, with a focus on the overall, superior sustainability score.

In terms of a superior score, there were several options to choose from. Recall from the literature review how the following metrics have been studied earlier:

- Morningstar Sustainability Rating (globe ratings)
- Portfolio ESG Score
- Portfolio Sustainability Score

The globe-ratings rate funds by comparing them to their category peers. Funds are listed from best to worst within their category and awarded “globes” based on their ranking as described in section 3.2. This implies that, in theory, a fund which currently sits on the highest possible rating (5 globes), can be given the worst possible rating (1 globe) when assigned to a different category. A score of “high”, or 5 globes indicates that a fund is in the top 10 percentile within its fund category. However, this category might have low average (and low high end) scores compared to other categories. As such, the MSR is best used when your investment options are confined to a specific category, e.g. “Europe Equity Large Cap.”

The Portfolio ESG Score and the Portfolio Sustainability Score are fairly similar in that both are created to work as a cross-category scoring system, making judgement on the portfolios’ underlying assets. A score of 50 is interpreted as: “The weighted-average score of the underlying companies, when compared to their industry peers is 50.” The interpretation is consistent, regardless of fund category. The difference between the two scores is that the sustainability score encompasses measurements of controversy. While there are arguments to be made for extracting these from the analysis, e.g. isolating ESG performance and achieving higher persistence in scores²¹, I see the benefit of adding controversy scores to be greater. Including the controversy score gives a more realistic picture of a fund’s sustainability performance. The score deductions are generally not that impactful on overall sustainability score as a function of the hurricane distribution. However, when the controversy deductions are impactful despite this type of distribution, it is imperative that investors are aware of it, as

²¹ Changes in controversy scores can affect persistence as they take place as a function of controversial incidents and has the potential (in severe cases) to affect the superior score in a larger degree than other sub-scores.

can be deduced by Nofsinger and Varma (2014)²². Furthermore, the isolated ESG performance is captured by pillar scores.

As mentioned in section 3.3, the current scoring system used to derive MSR is the Historical Portfolio Sustainability Score. Because this is newly implemented (October 2018), I have not found any research that analyses fund performance in relation to this metric yet. For this reason, and following the reasoning above, this study uses the Historical Portfolio Sustainability Score as the superior sustainability metric.

The other sustainability metrics used in this study are all disaggregated from the overall sustainability score. These comprise of the controversy score and the three pillar scores; an environmental, a social and a governance score.

6.4 Persistence in the data

As repeatedly mentioned, the intention of this study is to analyze how MSR affects mutual fund performance. In order to draw more robust conclusions from the analyses, the sample period was set to five years. While it would be interesting to analyze the relationship between mutual fund performance and ESG-ratings in periods of market turmoil, e.g. the financial crisis of 2007-2009, MSR was introduced in 2016 and does not contain the time-series properties one would need for such studies. Furthermore, because the dataset used in this study solely contains currently operating funds, survivorship bias would weigh too heavy to draw reliable conclusions. Here, I discuss some of these issues in more detail and argue why limiting the sample to five years provides more reliable results.

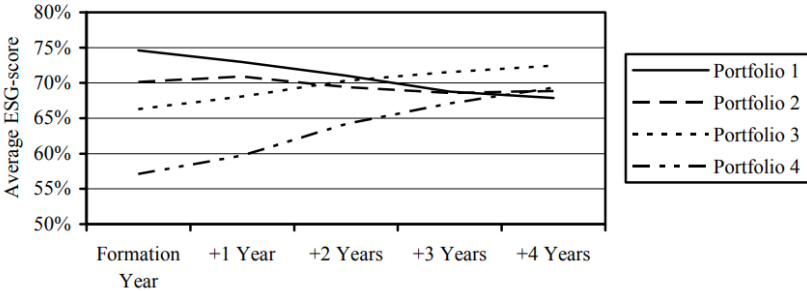
Persistence of sustainability ratings

Due to the cross-sectional nature of sustainability scores in the data set, I wanted to define a period for which it is reasonable to believe that the relationship between fund returns and sustainability scores persists. Wimmer (2013) studies the persistence of ESG scores in mutual funds. In the study, Wimmer gathers ESG ratings on a company level and recreates SRI-funds as listed by US SIF. In a similar fashion to Morningstar, the author excludes funds where less than 70% of the securities are covered by ESG data. Dividing the funds into quartiles based on their ESG scores, he finds that high ESG scores generally persist for two years on average

²² Exemplified in 4.4, Nofsinger and Varma (2014) illustrate how large controversies/scandals can lead to long term performance loss that is underestimated by the market.

and drop considerably afterwards. Contrarily, low rated funds gain momentum around the two-year mark and eventually catches up with the high rated group.

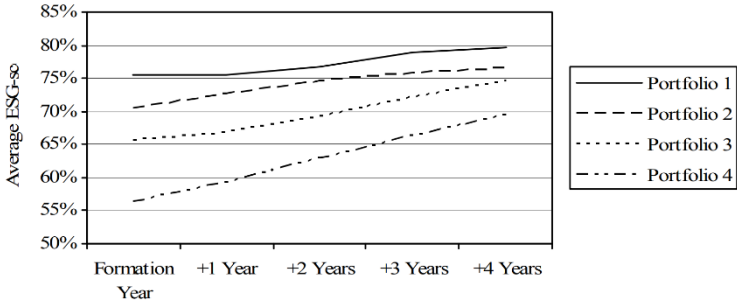
Figure 6.1. Wimmer (2013) Persistence of ESG ratings



Source: Wimmer (2013)

In essence, this means that investors cannot necessarily rely upon long-term continuation of ESG-scores, as scores seemingly converge towards a mean around the four-year mark. Wimmer attributes the convergence in scores to changes in portfolios’ holdings. When keeping the portfolio holdings fixed, he instead finds that ESG scores persist for at least four years. Consequently, he claims that there is a lack of appreciation of sustainability among fund managers.

Figure 6.2. Persistence of ESG ratings with constant portfolio holdings



Source: Wimmer (2013)

While Wimmer (2013) provides an indication of persistence in ESG ratings, the rating methodology Morningstar uses (for the Historical Portfolio Sustainability Score) is different from what Wimmer used in his study. First, it consists of a 12-month history of portfolio holdings. This will increase the period for which results are valid. Secondly, Morningstar created a methodology intended to align with portfolio analysis specifically, while Wimmer

simply applies the raw company (ASSET4) ESG scores to construct his portfolio scores. It is conceivable that this might also influence persistence.

In their thesis, Auran and Kristiansen (2016) argue that more narrow distributions will increase robustness of the sustainability scores, which seems reasonable to believe.

Additionally, data used by Wimmer (2013) covered a period from 2003 to 2009 and I would argue that fund managers on average are a lot more devoted to sustainability now, 10-15 years later, especially in the upper percentiles. As mentioned, Wimmer argues that a negligence of sustainability by fund managers seems to cause considerably higher fluctuations in scores.

I recognize the inexact science and that variable persistence might have implications for the validity of regression results. However, based on Wimmer (2013) and adjustments from the arguments above, this study uses quintiles in sustainability rankings and assumes persistence between the top- and bottom 20% categories and fund performance for five years.

Persistence of mutual fund performance

The data sample of fund returns used in this study exclusively includes funds that are operational as of December 2018. Funds that have been liquidated or merged with other funds prior to this date are as such not included in the sample. This creates a sampling bias known as survivorship bias. In essence, survivorship bias is problematic in research as funds that disappear tend to do so due to poor performance (Elton et al, 1996). If poor performers close up shop, samples that do not correct for attrition risk overestimating the expected return and underestimating associated risk. Furthermore, you run the risk that some of the variables studied correlate with attrition, creating spurious results. One can argue that the controversy score in this study is a likely example of a variable correlating with attrition. Controversies lead to high controversy deductions, and as Nofsinger and Varma (2014) exemplifies, controversies are often a cause of poor performance (and poor performance equates to higher chance of attrition).

According to Carhart et al (2002), this type of sampling bias affects almost every study on mutual fund performance. In their study on mutual fund survivorship they estimate the annual survivorship effect to approximately 1% of risk-adjusted performance for samples longer than 15 years. Similarly, Elton et al (1996) estimate a bias-effect of 0,97% in a 20-year sample. However, survivorship bias is inversely related to sample length. For one-year samples, the annual bias on risk-adjusted performance is estimated to 0,07% (7 basis point alpha) by

Carhart et al (2002). Elton et al (1996) calculate an annual 10-year sample survivorship bias of 0,40%.

Survivorship is a more nuanced variable than presented here. For example, management style can play a role. Realized returns can induce apparent persistence in performance (Brown et al, 1992). Loading on value and size plays a role in fund survival, as larger capitalization, value-based funds tend to outlive other, usually more aggressive investment styles (Carhart et al, 2002).

Limited by currently existing data, I was not able to account for survivorship biases in the study. It is regrettable that this sample bias cannot be accounted for, as idiosyncratic risk factors partake in causing attrition and one would expect these risk factors to be more present in low sustainability funds, as discussed in section 5.

Using a 5-year sample period, the expected effects of survivorship bias are likely no more than 0,2-0,3% of annual risk-adjusted returns, based on a linear aggregation of studies referenced above. However, the overall sample bias is not necessarily representative for all sections of the sample. For instance, small-cap funds are more exposed to survivorship bias than large-cap funds. Similarly, funds weighted towards growth stocks are more prone to bias from not including dead funds than value-weighted funds. In cases of a negative significant HML or positive significant SMB factors in the multi-factor regression model, these must be interpreted with caution as there might be an implicit additional risk attached. Survivorship bias is a thematic limitation in mutual fund performance analysis. It does not discount the validity of this study, but it should be taken into consideration when attempting to draw conclusions from regression results.

6.5 Descriptive statistics

In this section I present data and statistics on which I base the analyses in section 7.

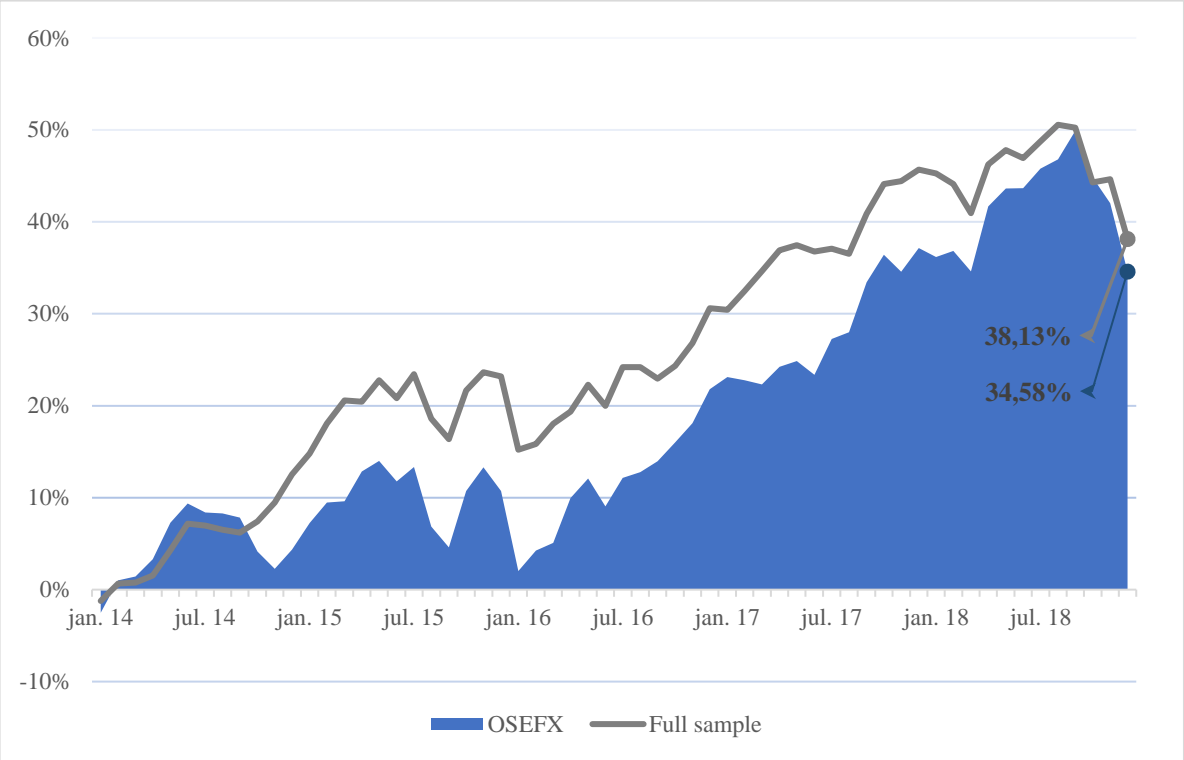
Throughout the section I reference tables placed in the appendix at the end of this paper.

Return series

For the full sample of 146 funds, the cumulative five-year simple returns of the equally weighted portfolio are 38,13%. The annualized simple returns are 7,63% with a standard deviation of 9,10%. The market beta for the full sample is 0,71. The benchmark, OSEFX has

annualized returns of 6,92% with a standard deviation of 10,7%. The cumulative returns for the portfolios are portrayed in figure 6.3.

Figure 6.3. 5-Year cumulative returns for OSEFX and the full, equally-weighted 146-fund sample.



Displayed in figure 6.3 are the cumulative returns for the equally weighted portfolio for the 146-fund sample and the OSEFX.

The return series seem to correlate for the period, except for the fall of 2014 when the oil crisis struck. Logically, the OSEFX is more invested in the oil market than the average fund in the full sample which partially consists of emerging market funds, sector specific funds and globally invested funds. The monthly returns for the five-year period have a .82 correlation. Ignoring the 8-month period between August 2014 and May 2015, the correlation is .88.

For all sustainability metrics, there is crafted a portfolio for each of the top 20% (High) and bottom 20% (Low) sustainability funds.²³ Portfolios created for this study are all equally weighted. While a value-weighting or a top/bottom heavy weighting might be more appropriate for certain analyses, equally weighting is kept consistent for the purpose of this

²³ For the remainder of this paper, “high sustainability”, “top quintile” and “High” (capital H) are interchangeable terms, referring to the top 20% sustainability portfolios. Likewise, “low sustainability”, “bottom quintile” and “Low” (capital L) refers to the bottom 20% sustainability portfolios.

study. It is worth noting that the controversy portfolios are ranked by sustainability rather than score. That is to say, the High portfolio includes the 20% lowest controversy scores, i.e. top 20% in terms of sustainability. Conversely, the Low portfolio consists of the highest controversy scoring funds. For all other portfolios, the description of high or low is consistent with the numerical value of the sustainability score. Descriptive statistics for the High and Low portfolios for each of the sustainability metrics are presented in table 6.2.

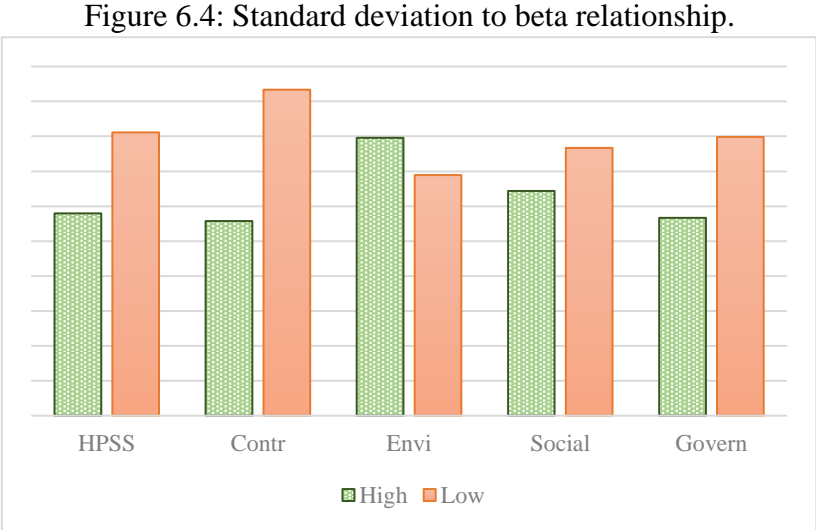
Table 6.2. Descriptive statistics of top and bottom quintile sustainability portfolios.

Sustainability Performance Statistics	<i>Portfolio</i>	<i>Returns Ann.(%)</i>	<i>Std.dev. Ann.(%)</i>	β	<i>Sharpe Ratio</i>
<i>HPSS</i>	High	7,35	9,51	0,82	0,77
	Low	7,04	10,06	0,62	0,70
	Diff	0,31	-0,55	0,20	-0,07
<i>Contr. Score</i>	High	6,92	10,15	0,91	0,68
	Low	9,17	10,83	0,58	0,85
	Diff	-2,25	-0,68	0,33	-0,17
<i>Environment</i>	High	7,93	10,03	0,63	0,79
	Low	6,30	9,51	0,69	0,66
	Diff	1,63	0,52	-0,06	0,13
<i>Social</i>	High	6,63	9,66	0,75	0,69
	Low	6,63	9,97	0,65	0,67
	Diff	0,00	-0,31	0,10	0,02
<i>Governance</i>	High	7,22	9,52	0,84	0,76
	Low	7,30	10,22	0,64	0,71
	Diff	-0,08	-0,71	0,20	0,05

In table 6.2 are the returns, standard deviations, betas and Sharpe Ratios for the top and bottom quintile sustainability portfolios for a variety of sustainability metrics. A positive (negative) figure in the diff row indicated that the high (low) sustainability portfolio has a higher (lower) reported figure than the low (high) portfolio.

The constructed High and Low sustainability portfolios' performance is in line with the market performance as no estimated t -or F-values provide statistical significance at the 10% level. Internally, within metrics, returns are similar between the top and bottom scoring portfolios, with two outliers. First, the High environment portfolio has realized 1,63 percentage points higher returns than the opposing Low portfolio, however the difference is not statistically significant. The second, and larger outlier lies with the Low controversy group, where the Low group outperformed the corresponding High group in terms of annualized returns with 2,25 percentage points, but again without statistical significance. The

Low controversy portfolio also carries the lowest market beta out of any portfolio created, but the highest standard deviation. A pattern where total variance is higher, while betas are lower is a recurring theme when comparing the low sustainability groups to their corresponding high sustainability groups. In the graph below, I plot the quotients for the portfolio standard deviations relative to the corresponding betas.



In figure 6.4, columns show how the high and low sustainability portfolios' standard deviations relate to their respective beta (calculated as standard deviation divided by beta). From left to right are the Historical Portfolio Sustainability Score, Controversy Score, Environ Score, Social Score and Governance Score portfolios.

The low sustainability portfolios carry a higher level of total risk compared to corresponding betas, with one exception being the environmental factor. Generally, a high standard deviation to low beta relationship implies higher idiosyncratic risk connected to the portfolio. However, as all betas are lower than 1, it is also possible that the index used as a market proxy in this study is not well suited to explain the variance in the sample returns.

Delving further into top and bottom quintile HPSS portfolios, I plot the monthly returns in the figure below.

Figure 6.5. Top and bottom quintile HPSS monthly returns.

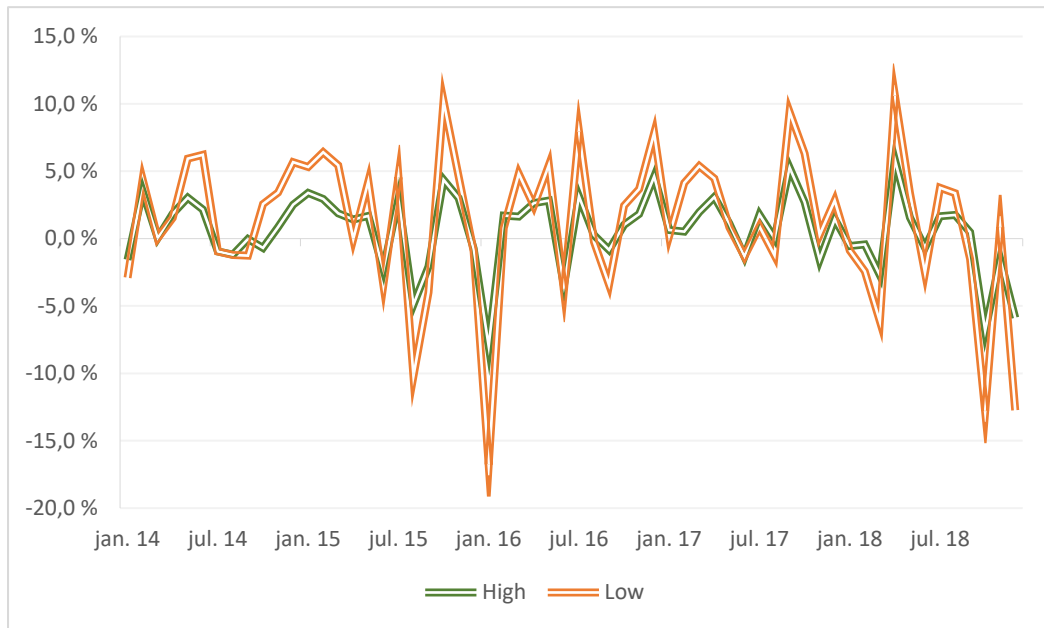
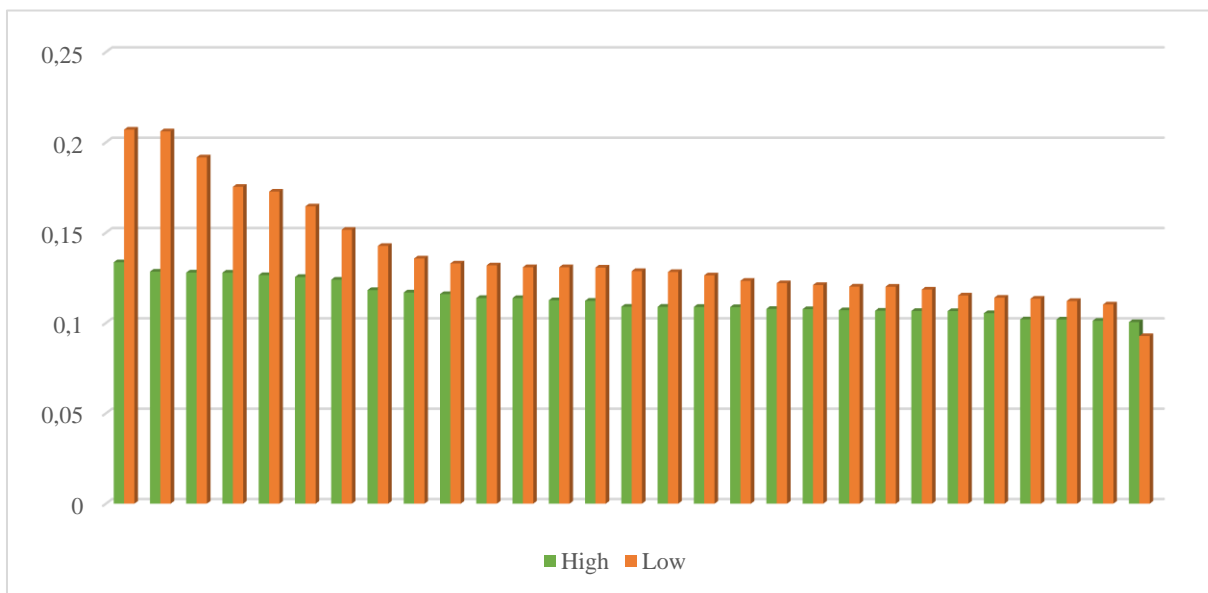


Figure 6.5 shows the monthly returns for the top and bottom quintile HPSS portfolios for the period between January 2014 and December 2018.

The first thing I notice is how the Low group is consistently more volatile for the entire period. Peaks as well as dips are more pronounced in the Low portfolio. To understand why, I disaggregate the portfolios and look at individual fund standard deviations.

Figure 6.6. Top and bottom quintile HPSS annualized standard deviations.

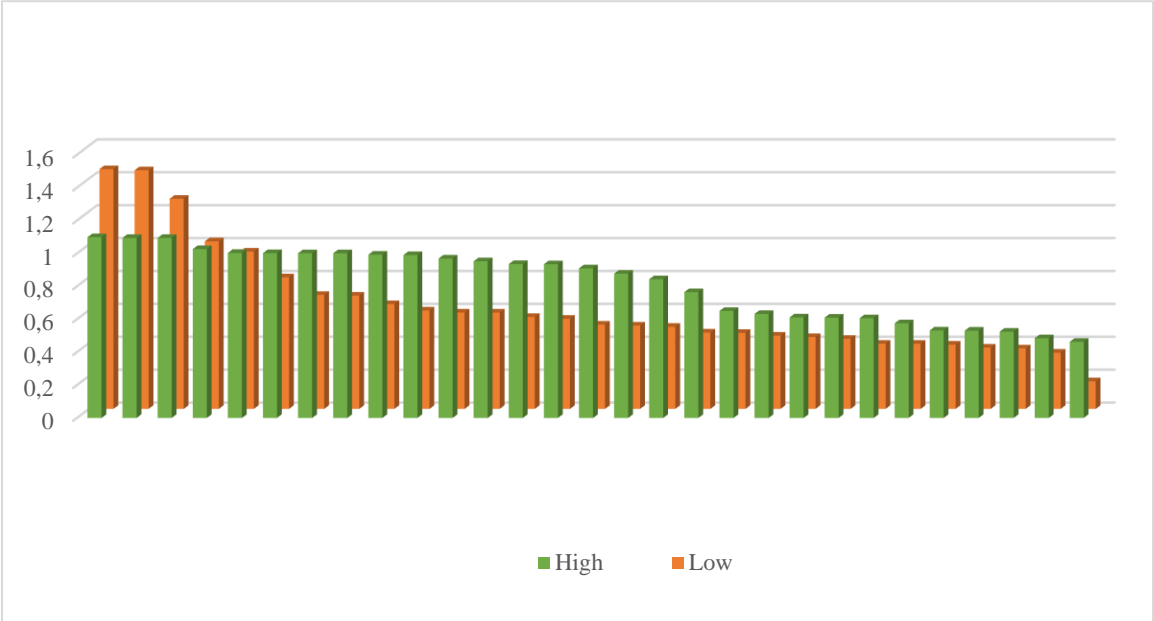


In figure 6.6, the top and bottom quintile HPSS portfolios are sorted from highest to lowest standard deviations within their groups. Each group consists of 29 funds.

From purely visual analysis it is evident how little standard deviations vary in the high sustainability group. Comparatively, the low sustainability group have drastically varying and generally higher standard deviations. The standard deviation of standard deviations in the High and Low group are 0,009 and 0,028 respectively. It is also worth mentioning that there are no less than 9 (roughly 1/3) observations in the Low portfolio with a higher standard deviation than the highest observation in the High portfolio. For an investor investing in a single fund, rather than a diversified portfolio of funds, the reported differences here could be of interest.

Comparing market betas, the Low portfolio has a .2 lower beta than the High portfolio. The betas for the funds are plotted in the tables below.

Figure 6.7. Top and bottom quintile HPSS portfolio beta values.



In figure 6.7, the top and bottom quintile HPSS portfolios are sorted from highest to lowest beta values (with benchmark OSEFX) within their groups. Each group consists of 29 funds.

Along the x-axis are the funds in order from highest to lowest beta values within the groups. First, due to the properties of the beta value (where at least a share is represented in the standard deviation), the graphs show similar patterns to the standard deviations, where the High group carries a consistency that is not reflected in the Low group. Over half of the funds in the high HPSS portfolio have a beta above .8, while only 5 funds in the Low portfolio can say the same. In the Low portfolio most betas are below .6, however, the 3 largest betas are all found in this group. Based on the figures above in a vacuum, it would seem like low sustainability score is connected to a polarizing management of risk; either high or low systematic risk.

The juxtaposition of the variation measures above provide insight to structural portfolio differences. There are seemingly systematic differences in how portfolios are managed at the high and low end of the sustainability spectrum. However, as mentioned earlier, low beta values could also be a function of a benchmark index that is not representing the sample well. The next section provides further insight.

Fund category allocation

Historically, sustainability funds have tended to invest more in large capitalization stocks than the market otherwise. At first glance, this seems to be the case for the data sample in this study as well, though not by a large margin. In Table A1 in the appendix, a similar table to table 6.1 is presented, but for all High and Low portfolios, showing portfolio holdings by category. For the full sample of 146 funds, 44,5% are placed in large-cap categories. The average amount of large-cap categorized funds in the High and Low groups is 53,1% and 39,2% respectively. Compared to the overall sample, the High groups seem to carry a slight, but present tilting towards size stocks. The opposite is true for the Low groups.

However, when defining large-cap solely as the “Europe Equity Large-Cap” category, the top and bottom sustainability quintiles instead average 49,7% and 8,3% respectively. Compared to the overall sample contribution of this category of 17,1%, this difference is significant. Derived from table A1 are the following statistics of fund categories in the top and bottom quintile sustainability portfolios:

Table 6.3. Percentage of allocation in large-cap categorized funds and in Europe.

		Large-cap %	Europe Equity Large-cap %	Europe %*
<i>HPSS</i>	High	58,6 %	55,2 %	96,6 %
	Low	41,4 %	0,0 %	6,9 %
<i>Contr</i>	High	3,4 %	0,0 %	86,2 %
	Low	65,5 %	13,8 %	13,8 %
<i>Environ.</i>	High	72,4 %	62,1 %	75,9 %
	Low	20,7 %	0,0 %	24,1 %
<i>Social</i>	High	72,4 %	72,4 %	100,0 %
	Low	31,0 %	0,0 %	10,3 %
<i>Govern.</i>	High	58,6 %	58,6 %	100,0 %
	Low	37,9 %	27,6 %	6,9 %

Table 6.3 shows the portfolio allocation by large-capitalization and European investment area.
 * In this column it is reported how much of the portfolios are invested in either Europe Equity Large-Cap or Europe Equity Mid/Small-Cap as categorized by Morningstar.

As seen from the table above, when isolating European categories, the distribution of large cap funds in the low sustainability groups essentially dissipates. In contrast, most large-cap allocation in the High groups is kept consistent when isolating for European markets. It seems as if global (ex. Europe) large-cap funds have a negative contribution to sustainability scores in this sample. This observation is interesting for several reasons. The consistency of high sustainability funds investing in Europe could broaden and simplify sustainability-driven investors' investment options. Doyle (2018) suggests that such differences is an issue of geographical bias, or differences in disclosure practices (leading to geographical bias). Whether or not that is the case here, a negative contribution to sustainability scores of non-European categorized funds is undoubtedly present and must be considered in further analyses.

Furthermore, when isolating European-categorized funds, it becomes apparent that there exists a large-cap weighing in high ranked funds. This supports earlier findings of Johnsen and Gjørlberg (2008), Dolvin et al (2017) and others where self-proclaimed sustainability funds carry a tilting towards size stocks. That is to say, funds with a self-proclaimed sustainability mandate show similar properties of high scoring funds in terms of capitalization.

It is worth noting that the controversy portfolios yet again act like outliers. The high sustainability controversy group is almost solely consisting of Europe Equity Mid/Small-Cap funds and includes none of the European large-cap funds. This is in contrast to all other high sustainability portfolios created. Is the chance simply lower for mid -and small capitalization funds to be deemed controversial? If so, why? Due to less extensive business practices, i.e. less involvement with controversial companies? Or maybe less visibility in traceable sources? Recall from section 3.2 how controversy scores are set by Sustainalytics' analysts based on news screening. A further notion of geographical bias is found in the low sustainability controversy group, where 65,5% of the portfolio consists of large-cap funds, but only a fraction of these are invested in Europe. This is interpreted as most controversial assets (as deemed by Sustainalytics) being global (ex. Europe) large capitalization companies.

In the previous sub-section, I illustrated the differences in standard deviations and betas between high and low HPSS portfolios. Based on the table above, it is conceivable that the geographical investment area has greater explanatory power for differences in return variance compared to that of the sustainability score. Possibly, there is a combination of the two variables. I investigate the relationship between geographic investment area, sustainability

and performance further in chapter 7.

Sustainability scores

Descriptive statistics for sustainability scores are presented in table A2 in the appendix.

Dolvin et al (2018) report an average Portfolio Sustainability Score for 43,98 in their sample of U.S. domiciled mutual funds. Comparatively, my full sample average HPSS is 51,83. To the untrained eye, a 7,85-score difference might not seem like much, but this difference is highly significant as derived from section 3.2. When isolating for European categories, the average is 55,55, possibly supporting the hypothesis of an existing geographical bias.

The difference in variation in the top and bottom quintile's HPSS figures are drastic, with a standard deviation of .85 in the high group and 1.36 in the low group. The standard deviations of the two samples are statistically different at the 1% level of significance. This could tell us that scores in the high end are more devoted to their level of sustainability and are reaching some sort of an asymptote. However, at the bottom percentiles, the standard deviations might be imprecise as a variation measure, due to greater impact of controversy scores.

For the controversy scores, the variance is not comparable, as the scores are awarded based on a hurricane scale, where severity increases exponentially with score. While the calculated standard deviations are vastly different, controversy scores do not follow a standard distribution, rendering the standard deviation meaningless as a measure of variation. It is expected that bottom percentiles (worst in terms of sustainability score contribution, i.e. highest scores) are quite distinguishable from the following percentiles, as they are much rarer to come by. Consequently, you will see many more funds with low scores than high scores.

For each of the pillar scores, the difference between the maximum and minimum observations are higher in the low scoring groups, with the high group having a statistically significant lower standard deviation at the 5% level for the environmental and governance factor. The difference in variation for the social factor does not carry statistical significance. Generally, this could be interpreted as the top end sustainability funds being more committed to their level of sustainability.

“ESG momentum”

For the purpose of this study, I have crafted a metric based on MSR, which I have named “ESG momentum.” The score is defined as follows:

$$ESGmom = Portfolio\ Sustainability\ Score - Historical\ Portfolio\ Sustainability\ Score$$

By subtracting the HPSS from the current score, I get a measure that tells me in which direction the sustainability of a fund has moved in the past year. A positive sign signifies an increasing score, while a negative sign signifies a decreasing score. As can be derived from the formula for HPSS in section 3.3, the HPSS is an average weighing of the last 12-months portfolio holdings, with more emphasis on recent holdings. This implies that the ESG momentum values recent months’ portfolio changes more than changes from early in the year. Descriptive statistics for crafted portfolios are presented in the table below.

Table 6.4. ESG momentum portfolio performance and sustainability statistics.

Portfolio performance statistics					Sustainability statistics			
Portfolio	Return Ann(%)	Std.dev. Ann(%)	β	Sharpe Ratio	Full sample	High	Low	
High	8,93	9,03	0,74	0,99	Mean	-0,15	0,7	-0,98
Low	7,43	9,20	0,69	0,81	Median	-0,14	0,49	-0,73
					Std.dev.	0,68	0,7	0,59
					Max.	3,08	3,08	-0,52
					Min.	-3,03	0,18	-3,03

The High portfolio realizes 3,3 percentage points higher annual returns when compared to the full sample, with a corresponding 7 basis points lower standard deviation. Consequently, the high scoring ESG momentum portfolio creates a higher Sharpe ratio than any other portfolio in this study, including the benchmark which has a .34 lower Sharpe. Purely based on statistics presented above, the Low portfolio does not stand out. Further analysis will provide deeper insight to the possibility of an “ESG momentum”-effect.

7. Research methodology, econometrics, results and analysis

The descriptive statistics paint a picture of how risk and returns scale with sustainability ratings. However, there are statistical implications that are not appropriately captured by a simple comparison of returns and standard deviations. In this chapter I use a variation of regression models to delve deeper into the performance of mutual funds at differing sustainability levels.

7.1 Fama-French three-factor regression

To better capture risk-adjusted performance, I adjust for common risk factors with the Fama-French three-factor regression model. To create the model, I use my 146 fund-sample of monthly return data for the 5-year (60-month) period between January 2014 and December 2018. Sorting the funds in to quintiles by sustainability score, I extract the top and bottom 20% funds for a variety of sustainability metrics in order to analyze the effect of each metric on the funds' performance. The resulting top quintile (High) and bottom quintile (Low) sustainability portfolios are equally weighted with $n=29$ funds. The superior sustainability metric in the analysis is the Historical Portfolio Sustainability Score (HPSS). This score is disaggregated into a controversy score and three pillar scores; an environmental, a social and a governance score.

The model is formulated as follows:

$$r_i - r_f = \alpha_i + \beta(r_m - r_f) + \gamma_1 SMB + \gamma_2 HML + \varepsilon_i$$

Where

r_i is the monthly equally weighted average sample returns for the funds in the sample,

r_f is the measure for risk-free rate, and

r_m is the monthly market returns.

The benchmark in the study is the Oslo Børs Mutual Fund Index (OSEFX) and the risk-free rate is a Norwegian 3-year sovereign bond. For the SMB and HML-factors I use U.S. market data, downloaded from Kenneth French' website.

The results are reported in the table below.

Table 7.1. Fama-French three factor regression results.

<i>Morningstar sustainability metric</i>	<i>Portfolio</i>	<i>Alpha</i>	<i>β</i>	<i>SMB</i>	<i>HML</i>	<i>Adj. R²</i>
HPSS	High	0,001 (1,00)	0,81(§§§) (-4,38)	0,001(*) (1,95)	-0,001(**) (-2,26)	0,86
	Low	0,002 (0,88)	0,59(§§§) (-4,58)	0,003(**) (2,46)	-0,002 (-1,65)	0,49
	Diff	-0,001 (-0,52)	0,22(*) (3,27)	-0,002(**) (-2,02)	0,000 (0,73)	0,15
Contr. Score	High	0,001 (0,68)	0,90(§§§) (-3,03)	0,001(*) (1,82)	0,000 (0,43)	0,93
	Low	0,004 (1,29)	0,55(§§§) (-4,17)	0,002(*) (1,86)	-0,003(*) (-1,65)	0,37
	Diff	-0,004 (-1,11)	0,35(***) (3,33)	-0,002 (1,34)	0,003(**) (2,12)	0,18
Environment	High	0,003 (1,03)	0,62(§§§) (-4,44)	0,002(*) (1,93)	-0,003(***) (-2,74)	0,52
	Low	0,001 (0,66)	0,66(§§§) (-4,69)	0,002(**) (2,34)	-0,001(*) (-1,20)	0,62
	Diff	0,001 (0,71)	-0,04 (-0,75)	0,000 (-0,05)	-0,002 (-2,55)	0,07
Social	High	0,001 (0,53)	0,74(§§§) (-4,08)	0,001(*) (1,69)	-0,002(**) (-2,43)	0,72
	Low	0,002 (0,77)	0,62(§§§) (-4,46)	0,002(**) (2,37)	-0,001 (-1,05)	0,62
	Diff	-0,001 (-0,53)	0,12(**) (2,03)	-0,001 (-1,56)	-0,000 (-1,01)	0,01
Governance	High	0,001 (1,00)	0,83(§§§) (-4,52)	0,001(*) (1,91)	-0,001(*) (-2,00)	0,90
	Low	0,002 (0,89)	0,61(§§§) (-4,25)	0,002(**) (2,21)	-0,002 (-1,57)	0,49
	Diff	-0,001 (-0,57)	0,22(***) (2,84)	-0,002(*) (-1,69)	-0,001 (-0,88)	0,11

Reported in table 7.1 are the OLS estimates for all top and bottom quintile sustainability portfolios ($n=29$), as well as for the difference portfolios. Stars in the parentheses signify coefficient values different from 0 at a level of significance as follows: (*) = 10% level, (**) = 5% level, (***) = 1% level. Paragraph icons in the parentheses signify coefficient values different from 1 at a level of significance as follows: (§) = 10% level, (§§) = 5% level, (§§§) = 1% level. Numbers in the parentheses are the estimated t-values attached to the corresponding coefficient.

In the leftmost column in table 7.1 are the Morningstar sustainability metrics covered in chapter 3; The Historical Portfolio Sustainability Score, the Controversy Score and a score for each of the ESG pillars. In the second column are the portfolio sustainability criteria where: High = top 20%, Low = bottom 20% and Diff = the difference portfolio between the two aforementioned.

All alphas for the high and low sustainability portfolios carry a positive sign but are not statistically significant. For reference, the highest t-statistic connected to a three-factor alpha in the regression is with the low sustainability controversy group, where $t=1,29$. An implication of statistically insignificant alphas is that the performance of the portfolios is in line with the performance of the market proxy. Therefore, when adjusting for SMB and HML factors, I find no evidence of a statistically significant difference between the high and low sustainability portfolio alphas, implying that there currently exists no (statistically significant) financial benefits or drawbacks from investing in high (or low) ESG-rated funds.

As expected, top and bottom quintile sustainability portfolios are heavily affected by market returns. The conception of differences in market betas by sustainability, as portrayed in section 6,5 are affirmed through the regression results. Generally, the high sustainability portfolios are more susceptible to market risk than the corresponding low sustainability portfolios. That being said, large differences in adjusted R-square values indicate that the difference could be due to factors not captured by the model. Assuming that the OSEFX is the true benchmark of the sampled funds, a low R-square would indicate a higher level of idiosyncratic risk. However, as established in section 6.5, low sustainability portfolios in this sample tend to consist of funds investing outside of European markets, which is not the case for the benchmark portfolio OSEFX.

Several of the size and value factors are significant in varying directions and levels of significance. Generally, both the high and low sustainability portfolios seem to be influenced by growth -and small-cap stock returns. The difference-portfolios carry statistical significance in two cases for the SMB factor, where the low sustainability funds load more on small-cap than the high sustainability funds. For the HML-factor, only the controversy score difference-portfolio shows signs of factor loading. While some of the coefficients are significant from a statistical standpoint, the contribution of the SMB and HML-factor to the overall performance is minuscule. Furthermore, some portfolios, e.g. the top quintile environmental and social

portfolios have a large-cap allocation percentage of 72% (see table 6.3). It is unintuitive to interpret these as loading on small-cap returns, despite the regression results. The reason for this disconnect is likely a capitalization bias created by equally weighing of the portfolios, favoring smaller capitalization assets. The average diversified fund does not have its capitalization comprise of 29% small-cap assets, which is the allocation percentage in these data.

While the risk-adjusted performance of the created portfolios is comparable to the market, it is evident that investment area is an imperative factor in the distribution of sustainability ratings. To exemplify, I perform a simple binomial test, estimating the probability of the High HPSS group containing 28 or more European funds out of 29 total spots.

Figure 7.1. Binomial distribution of European funds (46% of sample) in the High HPSS portfolio.

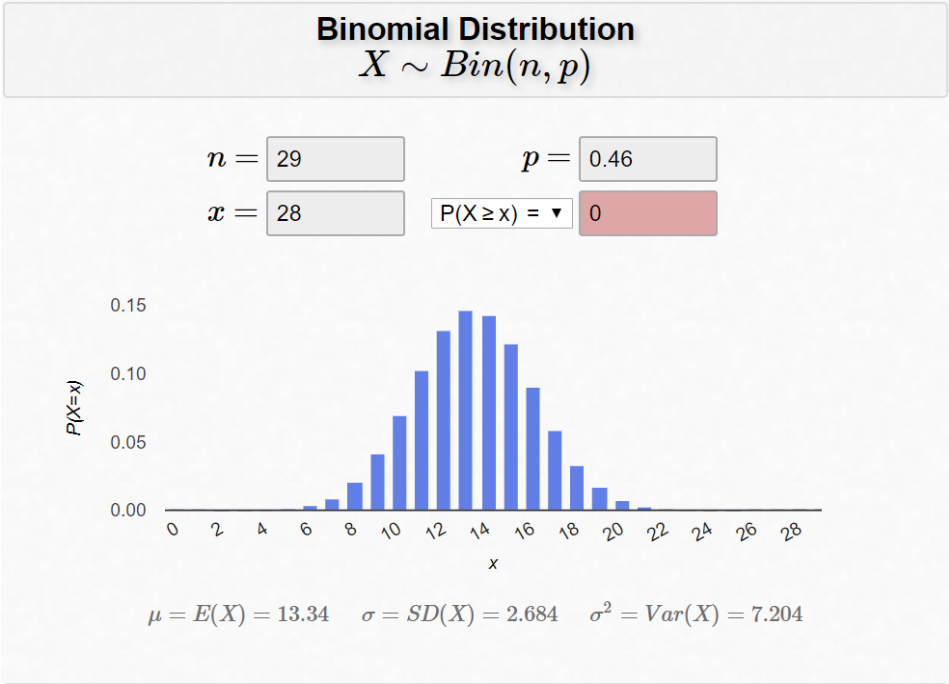


Figure 7.1 shows the binomial distribution of funds categorized as European appearing in a top or bottom quintile portfolio. European-categorized funds cover 46% (67 funds) of the 146-fund sample. The High HPSS portfolio consists of 28 European-categorized funds, and 1 non-European-categorized fund. The p-value shows the probability of this happening, given a bias-free distribution of sustainability scores.

The distribution displayed in figure 7.1 shows that the expected amount of European-categorized funds in a top or bottom quintile portfolio is 13. Derived from table 6.2, the average contribution of European-categorized funds is 92% (26,7 funds) in the high

sustainability portfolios and 12% (3,5 funds) in the low sustainability portfolios. The High HPSS portfolio comprises of 28/29 European-categorized funds. Given equal probability to receive a certain sustainability score across fund-categories, the chance of a created portfolio to consist of ≥ 28 European-categorized funds is 0,000%. This finding makes me question the robustness of results from the regression above as there appears to exist a geographical bias related to the distribution of sustainability scores. To test whether patterns in regression coefficients remain consistent based on sustainability ratings alone, I isolate for European-categorized funds and redo the regression.

European-categorized funds encompass 67 (46%) of the 146 funds sample. Dividing the funds into quintiles, the top and bottom quintile each consists of 13 funds. I apply the same regression model as above to the European-categorized top and bottom quintile sustainability portfolios. Results are reported in the table below.

Table 7.2. Fama-French three-factor regression results: European categories.

<i>Morningstar sustainability metric</i>	<i>Portfolio</i>	<i>Alpha</i>	<i>β</i>	<i>SMB</i>	<i>HML</i>	<i>Adj. R²</i>
HPSS	High	0,004(*) (1,77)	0,64(§§§) (-4,96)	0,002(**) (2,40)	-0,003(***) (-3,42)	0,62
	Low	0,002 (0,54)	0,75(§§§) (-4,34)	0,002(***) (2,57)	-0,001 (-1,17)	0,77
	Diff	0,003(***) (2,72)	-0,12(***) (-3,17)	-0,000 (-0,79)	-0,002(***) (-5,02)	0,37
Contr. Score	High	0,002 (0,96)	0,72(§§§) (-4,41)	0,002(**) (2,03)	-0,001 (-1,63)	0,71
	Low	0,000 (0,30)	0,82(§§§) (-3,49)	0,002(***) (3,14)	-0,000 (-1,12)	0,83
	Diff	0,001 (1,46)	-0,10(***) (-3,23)	-0,000 (-1,05)	-0,000 (-1,47)	0,18
Environment	High	0,001 (0,77)	0,77(§§§) (-4,00)	0,002(***) (3,31)	-0,001 (-1,54)	0,78
	Low	0,000 (0,10)	0,84(§§§) (-3,60)	0,001(***) (2,74)	-0,000 (-0,36)	0,86
	Diff	-0,000 (1,41)	-0,07 (-2,39)	0,000(**) (2,37)	-0,000(**) (-2,56)	0,18
Social	High	0,003(*) (1,69)	0,71(§§§) (-4,75)	0,002(***) (3,18)	-0,002(**) (-2,43)	0,74
	Low	0,002 (0,89)	0,72(§§§) (-4,30)	0,002(**) (2,37)	-0,000 (-0,51)	0,69
	Diff	0,001 (1,44)	-0,01 (-0,05)	0,000 (1,19)	-0,001(***) (-3,81)	0,17
Governance	High	0,003(*) (1,95)	0,72(§§§) (-5,19)	0,001(**) (2,16)	-0,002(***) (-3,22)	0,77
	Low	0,001 (0,75)	0,78(§§§) (-4,18)	0,002(***) (2,90)	-0,000 (-0,46)	0,82
	Diff	0,002(**) (2,10)	-0,06(**) (-2,11)	-0,000 (-0,96)	-0,002(***) (-4,67)	0,32

In table 7.2, I report the OLS estimates for all top and bottom quintile sustainability portfolios ($n=13$), as well as for the difference portfolios. Stars in the parentheses signify coefficient values different from 0 at a level of significance as follows: (*) = 10% level, (**) = 5% level, (***) = 1% level. Paragraph icons in the parentheses signify coefficient values different from 1 at a level of significance as follows: (§) = 10% level, (§§) = 5% level, (§§§) = 1% level. Numbers in the parentheses are the estimated t-values attached to the corresponding coefficient.

Isolating for European-categorized funds, I find evidence of market outperformance by the High HPSS portfolio, as well as the High social and High governance portfolios. The HPSS difference-alpha is positive and statistically significant at the 1% level. In accordance with previous research (Fulten et al, 2012; Han et al, 2016), the outperformance seems to stem from the governance-factor, as the governance-difference portfolio also conveys an alpha. With reservations for possible statistical uncertainties, a 0,3% monthly alpha difference between the High and Low HPSS is a significant finding, indicating a positive relationship between sustainability and excess risk-adjusted returns.

The 0,4% monthly HPSS alpha comes at the cost of increased idiosyncratic risk connected to the portfolio (i.e. lower appraisal ratio²⁴), as R-squared values are generally lower for the high sustainability portfolios. Removing idiosyncratic risk through diversification usually carries a transaction cost to the investor, a cost that is not captured by the model. However, these costs do not outweigh an alpha of this magnitude.

A pattern of lower R-squared values connected to High portfolios is dissimilar from what is found in the full-sample regression. Also different is the pattern observed in the market factor. Contrary to the full-sample regression, high sustainability portfolios are less susceptible to market risk when isolating for a European investment area. A general interpretation of this beta to R-square relationship is that, when competing in the same market, higher sustainability provides the investor with lower market risk and higher unsystematic risk. Looking past the numbers however, it appears that the high sustainability portfolios carry a higher fraction of non-Norwegian (geographical investment area) Nordic funds, while the Low portfolios are mostly invested in Norway. Again, a notion of geographical bias presents itself in the distribution sustainability ratings. In this case it is likely arbitrary as the deviation between sustainability scores in the 67-fund European-categorized sample is very low.²⁵

The size and value-factors carry the same patterns as with the full-sample regression. In both High and Low portfolios there is a general loading on small-cap and growth-stock returns when adjusting for the American market factors. The difference between High and Low portfolios is generally not statistically significant for the SMB-factor and generally highly

²⁴ The appraisal ratio is a measure for excess performance to the “residual standard deviation,” i.e. what remains after subtracting market risk. $AR = \alpha / \sigma_p$

²⁵ The difference in sustainability score between the top ranked fund and the median is 3,9. Comparatively, this difference is 8,9 in the full sample. The low score-deviations creates a situation where it is more random which fund is placed in the top quintile. This is not true for the bottom quintile portfolio where scores have a wider range, as seen in table A2 in the appendix.

significant (negative) for the HML-factor, meaning that the High-portfolios load more on growth stock-returns than corresponding Low portfolios. While this pattern is consistent it is an inaccurate description of the portfolios' allocation, as discussed previously. In the High HPSS portfolio 10/13 funds are categorized as European Equity Large Cap. In the corresponding Low portfolio 12/13 are categorized as Europe Equity Mid/Small Cap. This indicates a bias towards large-cap funds in the distribution of sustainability scores.

7.2 ESG-momentum

The analysis in the previous section explains some of the performance variance by sustainability scores. But how does performance vary by historically changing scores? To answer this, I have crafted a metric for the purpose of measuring the momentum-effect of changing ESG-scores. The “ESG-momentum” is defined as:

$$ESGmom = Portfolio Sustainability Score - Historical Portfolio Sustainability Score$$

Similar to regressions in 7.1, high and low sustainability portfolios are equally weighted, consisting of 29 funds with 60 monthly observations from January 2014 to December 2018.

Table 7.3: Fama-French regression results: ESG-momentum.

ESG metric	Portfolio	Alpha	β	SMB	HML	Adj. R ²
ESG Momentum	High	0,003(**) (2,08)	0,73(\$\$\$) (-5,50)	0,002(**) (2,55)	-0,001(**) (-2,04)	0,80
	Low	0,002 (1,10)	0,67(\$\$\$) (-5,05)	0,002(***) (2,14)	-0,002 (-2,01)	0,66

In table 7.3, I report the OLS estimates for all top and bottom quintile sustainability portfolios (n=29). Stars in the parentheses signify coefficient values different from 0 at a level of significance as follows: () = 10% level, (**) = 5% level, (***) = 1% level. Paragraph icons in the parentheses signify coefficient values different from 1 at a level of significance as follows: (\$) = 10% level, (\$\$) = 5% level, (\$\$\$) = 1% level. Numbers in the parentheses are the estimated t-values corresponding to the related coefficient.*

The top quintile ESG-momentum portfolio creates an alpha significant at the 5% level. From an investor’s perspective, this result could have several possible implications. Performance gain from increasing ESG-scores could indicate an existing reward connected to buying low sustainability funds and investing efforts into improving the fund’s sustainability. If one

believes that these numbers carry practical accuracy, firm owners can increase their risk-adjusted performance by 3,6 percentage points yearly by creating an ESG-momentum score of 1 – a task that should be attainable for low rated firms. While the ESG-momentum carries diminishing returns, i.e. you can not expect to sustain a 1-point ESG-momentum for many consecutive periods, even a one-time performance gain of this magnitude would be astonishing.

Another aspect is provided by Wimmer (2013) who attributes changes in ESG-scores to changes in portfolio holdings. If true, it is possible that swapping out low sustainability funds in favor of high sustainability funds is beneficial for increasing risk-adjusted performance. However, given that alphas in table 7.1 do not carry statistical significance, Wimmer's hypothesis is not applicable here. That is to say, high sustainability funds in the full sample are not out-performing low sustainability funds.

The portfolio with the 20% lowest ESG-momentum scoring funds shows no signs of gaining performance nor suffering performance loss from dropping in sustainability rating. In essence, this means that funds are not punished for reducing their sustainability performance but rewarded for increasing it.

7.3 Cross-sectional analysis

To further analyze the relationship between fund performance and Morningstar Sustainability Ratings, I supplement the study with a series of cross-sectional analyses. Throughout this section I will add and subtract explanatory variables to analyze the sensitivity of average returns in the data with respect to betas and sustainability ratings. To establish a baseline, I first look at how average returns change with changes in the market betas as estimated when employing OSEFX as the benchmark index. I gather annualized average excess returns for all 146 funds for the five-year sample and generate OLS estimates with fund betas as the independent variable:

$$\bar{r}_i = \alpha_i + \gamma_1 \beta_i + \varepsilon_i$$

Where,

\bar{r}_i is the average return for the fund i during the period January 2014 to December 2018, and β_i is the beta of fund i for the same period.

The rationale behind this is to examine how the beta coefficient varies with the addition of other explanatory variables. The regression results are presented in table 7.4.

Table 7.4. Cross-sectional regression results: Average returns with respect to beta values.

<i>Regression statistics</i>			
Adj. R-square		0,21	
Standard error		0,04	
Observations		146	
	<i>Coefficients</i>	<i>Standard error</i>	<i>t-Stat</i>
Intersection	0,132***	0,9 %	14,1
Beta	-0,078***	0,1 %	-6,27

Table 7.4 shows the regression results for an analysis with average returns of n=146 Norwegian ISIN funds with respect to their beta as estimated with OSEFX as benchmark.

Stars signify a coefficient value statistically significant as follows: () = 10% level, (**) = 5% level, (***) = 1% level.*

- When beta increases by 1%, average returns **decrease** by 0,078%.

Despite the use of market betas as an independent variable, the model has relatively low explanatory power, with an adjusted R-square of 0,21. Strangely, betas are negatively correlated with returns. This is unintuitive and is possibly a result of the benchmark of choice not being representative for which indices the funds in the study benchmark themselves against. Moreover, it could imply that the fund returns are tied to systematic factors not included in beta. As a quick mention, MSR is not applied to securities that are using a shorting strategy. Recall from table 7.1 how the top sustainability quintile contains a disproportionate amount of European categorized funds. For reference, when performing the analysis based on the top and bottom HPSS quintile, the betas are -0,04 and -0,15 respectively, both statistically significant at the 1% level. I study the presence of a non-positive beta further below.

The next variable I add to the analysis is the Historical Portfolio Sustainability Score to look at whether changes in ESG metrics have explanatory value for changes in fund returns. I formulate the model as follows:

$$\bar{r}_i = \alpha_i + \gamma_1\beta_i + \gamma_2HPSS_i + \varepsilon_i$$

Where,

$HPSS_i$ is the Historical Portfolio Sustainability Score of fund i for the same period.

Results are presented in table 7.5.

Table 7.5. Cross-sectional regression results: Average returns with respect to beta and HPSS.

<i>Regression statistics</i>			
Adj. R-square		0,23	
Standard error		0,04	
Observations		146	
	<i>Coefficients</i>	<i>Standard error</i>	<i>t-Stat</i>
Intersection	0,059	3,6 %	1,63
Beta	-0,089***	1,3 %	-6,65
HPSS	0,002**	0,1 %	2,08

Table 7.5 shows the regression results for an analysis with average returns of n=146 Norwegian ISIN funds with respect to their beta as estimated with OSEFX as benchmark and their Historical Portfolio Sustainability Score. Stars signify a coefficient value statistically significant as follows: () = 10% level, (**) = 5% level, (***) = 1% level.*

Both beta and HPSS coefficients show a statistically significant contribution to changes in average returns. The coefficients can be interpreted as follows:

- When beta increases by 1%, average returns **decrease** by 0,09%.
- When HPSS increases by 1, average returns **increase** by 0,2%.

Based on the regression results it would seem like buying the highest possible scoring fund would yield the best returns. While statistically significant, the coefficient is small at 0,2%.

Betas continue to be negatively correlated with returns. Theoretically, one would expect a positive relationship between average returns and betas. However, an absence of this relationship is a common problem distinguishing theory and empirical records. Fama and French (2004) studies this particular topic and find no evidence of a relationship between beta and average returns, despite the suggestions by CAPM. To cement the frequency of a result like this, Fama and French (2004) states: “We judge it unlikely that alternative proxies for the market portfolio will produce betas and a market premium that can explain average returns on

these portfolios.” This criticism of the CAPM, known as “the beta anomaly”²⁶ has been documented for decades, back to Fischer Black in the 70s (Swedroe, 2018).

To test whether betas are simply behaving unintuitively or could be a source of error due to the choice of benchmark, I repeat the analysis with a selected basket of funds that closer mirror the performance of OSEFX. In the basket are 46 funds where, according to Nordnet the majority of which are benchmarked against OSEFX. Some are benchmarked against OSEBX and a minority are tracking Norwegian markets but using a different benchmark²⁷ than the two aforementioned. To avoid repetitiveness, tables are placed in the appendix for the remainder of the analysis.

For the sub-sample of 46 funds, table A3 in the appendix can be summarized as follows:

- When beta increases by 1%, average returns **increase** by 0,003%.
- When HPSS increases by 1, average returns **decrease** by 0,03%.

When the sample is selected based on the funds’ benchmark, the beta essentially disappears as it is no longer statistically significant. Repeating my comments above, one would predict a positive relationship between average returns and betas, a relationship that is commonly missed in application. Regarding the HPSS, it is interesting to see that the returns negatively correlate with sustainability in this sub-sample. Yet again, it is evident that geography plays a part in how scores are distributed.

Throughout the last two chapters I have established the presence of a geographical bias in the distribution of sustainability ratings. Based on the results above, I hypothesize that returns gained from HPSS in table 7.5 are due to the geographical factors rather than sustainability scores. To test the hypothesis, I create a dummy variable to estimate the effects of having a European investment area. To further test how the HPSS and beta coefficients change, a set of dummy variables for the top and bottom quintile HPSS funds is added. The model is formulated as follows:

²⁶ “The beta anomaly” refers to the frequency in empirical findings where low beta portfolios provide superior performance.

²⁷ Examples of funds with a different benchmark are certain Danske Bank funds (e.g. Danske Invest Norge I or Danske Invest Norske Aksjer I) which primarily invests in Norwegian stocks but specifies FXLT or SXLS as their benchmarks.

$$\bar{r}_i = \alpha_i + \gamma_1\beta_i + \gamma_2HPSS + \gamma_3DummyEU_i + \gamma_4DummyTop20\%_i + \gamma_5DummyBot20\%_i + \varepsilon_i$$

Where,

DummyEU_i is a dummy variable for fund *i* where 1 = fund is placed in a Morningstar Global category defined as European and 0 = fund is placed in any other category,

DummyTop20%_i is a dummy variable for fund *i* where 1 = fund is in the top 20% HPSS portfolio from section 7.1 and 0 = fund is not in this portfolio, and

DummyBot20%_i is a dummy variable for fund *i* where 1 = fund is in the bottom 20% HPSS portfolio from section 7.1 and 0 = fund is not in this portfolio.

Results are presented in table A4 in the appendix. When adding the discussed variables, the HPSS is no longer found to positively impact average returns. Instead, the coefficient is small and negative, significant at the 5% level. A positive effect is found in the EU variable, indicating a positive relationship between average returns and Norwegian ISIN funds that invest in European markets. The top 20% HPSS dummy does not carry statistical significance, while the bottom 20% funds are found to provide lower returns in comparison to the rest of the sample.

Summarized, this indicates that in the sample, average returns are higher for European investment areas, while sustainability scores contribute to returns in a slight and negative manner. It appears that the negative contribution stems from the bottom quintile funds.

Lastly, I study the effect of ESG-momentum on fund returns. Using the ESG-momentum and fund betas as right-hand side variables, the model is formulated as follows:

$$\bar{r}_i = \alpha_i + \gamma_1\beta_i + \gamma_2ESGmom_i + \varepsilon_i$$

Where,

ESGmom_i is the ESG-momentum factor as described in section 6.5.

Results are reported in the table A5 in the appendix and can be summarized as follows:

- When beta increases by 1%, average returns **decrease** by 0,081%.
- When ESG-momentum increases by 1, average returns **increase** by 1,1%.

Betas continue to be negatively correlated with returns in a similar magnitude as in regressions above. The ESG-momentum score is found to positively impact fund returns, statistically significant at the 1% level. According to the model, one point of ESG-momentum is expected to increase average returns by 1,1%.

8. Summary and conclusion

Sustainable investing is a hot topic in the world of finance. Partly, this is due to increased ethical focus, but also a presumed financial gain linked to sustainable assets. Published numbers show that assets categorized as “sustainable investing”, “SRI”, “ESG” etc. are increasing rapidly on a global, year-to-year basis. Through this paper I have outlined some of the issues regarding these figures and the current state of “sustainable investing.”

Furthermore, I conducted a series of empirical analyses, which hopefully adds something to existing literature.

In the analysis, I adjust for common risk-factors with the Fama-French three-factor model. Employing the Morningstar Sustainability Rating system as a measure of sustainability I find no evidence of market outperformance by either the top or bottom quintile sustainability funds when analyzing a 146-fund sample of Norwegian ISIN funds. This initial finding implies no current rewards from screening for sustainability.

However, a geographical bias in the distribution of sustainability ratings likely causes some spurious results. When filtering the sample for a more homogenous investment area, defined by Morningstar Global categories as “European,” the top quintile sustainability portfolio is found to provide a 0,4% monthly alpha. The market outperformance is attributed to greater social and governance scores, while the environmental factor is not found to affect risk-adjusted returns in either direction. Furthermore, the top sustainability quintile outperforms the bottom quintile with a 0,3% monthly alpha, statistically significant at the 1% level. A positive relationship between sustainability and risk-adjusted performance is in line with findings by Kumar et al (2016), van Beurden and Gössling (2008) and others. Generally, the model shows that the high sustainability portfolios carry a larger portion of idiosyncratic risk, implying a lower appraisal ration. Removing the excess idiosyncratic risk comes with a cost to the investor that is not included in the model but is unlikely to negate an alpha of this magnitude.

While the fund returns in the data are found to tilt towards growth and small-cap stock returns, the top (bottom) quintile portfolios consistently contain a disproportionate amount of large-cap (small-cap) categorized funds. There is a likelihood of a capitalization bias in the distribution of sustainability ratings as suggested by Doyle (2018). High sustainability

equating large capitalization is not a revolutionary finding but supports findings by Johnsen and Gjørlberg (2008), Dolvin et al (2017), Auran and Kristiansen (2016) and others.

When subtracting the historical sustainability score from the current score I get a metric which tells me in which direction the funds' sustainability has moved the past 12 months. Employing the Fama-French three-factor model, the regression shows that the funds that have increased their rating the most are associated with a 0,3% monthly alpha, statistically significant at the 5% level. Seemingly, there exists an "ESG-momentum" effect in the data. This could indicate that it is in the shareholders' best interest to work towards improving sustainability.

Lastly, I study the relationship between sustainability ratings, betas and average fund returns in a series of cross-section analyses. The data suffers from what is known as the "beta-anomaly" – a concept that contradicts the CAPM as returns and betas have a non-positive relationship. This could partly be due to a poor choice of market proxy, but the non-positive relationship persists when adjusting the sample for the funds' specified benchmark. Initially, sustainability is found to positively correlate with average returns. However, when adding a geographical dummy to the analysis, the sustainability-related coefficients essentially dissipate (no longer provide positive impact), while European-categorized funds are outperforming other investment areas. An interpretation of this is that increased average fund returns cannot be attributed to high sustainability ratings, but rather the geographical investment area.

The top quintile ESG-momentum portfolio is found to provide 1,1% increased average returns with one point of ESG-momentum. Consistent with the Fama-French regression regarding ESG-momentum, performance gains seem to reside with increasing ESG-scores.

Due to the current state of ESG-ratings and lack of existing longitudinal data I was not able to account for survivorship bias in this study. Survivorship bias will conceptually affect low sustainability funds in a greater degree than high sustainability funds. Both because sustainability has presumed inherent downside risk protection, but also due to higher prevalence of large-cap assets in the top quintile portfolios. When available, I suggest that further studies examine the relationship between sustainability ratings and longer term, survivorship bias-free data.

9. Reference list

- Adler, T. & Kritzman, M. The Cost of Socially Responsible Investing. *Journal of Portfolio Management*, 35(1), p.52-56. Available at DOI: 10.3905/JPM.2008.35.1.52
- Auran, A. K & Kristiansen, A. (2016). *Investing in sustainability. The risk-adjusted performance of European mutual funds committed to sustainable and responsible investing*. [Master's Thesis]. Bergen: Norwegian School of Economics
- Bauer, R. & Derwall, J. & Hann, D., *Employee Relations and Credit Risk* (October 3, 2009). Available at SSRN: <https://ssrn.com/abstract=1483112> or DOI: <http://dx.doi.org/10.2139/ssrn.1483112>
- Bauer, R. & Koedijk, K. & Otten, R., (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7), p.1751-1767. Available at DOI: 10.1016/j.jbankfin.2004.06.035
- Beurden, P. & Gössling, T. (2008). The Worth of Values – A Literature Review on the Relation Between Corporate Social and Financial Performance. *Journal of Business Ethics*. 82. p.407-424. Available at DOI: 10.1007/s10551-008-9894-x
- Bloomberg (2017) *Global sustainable investments grow 25% to \$23 trillion* [Internet] Available at: <https://www.bloomberg.com/professional/blog/global-sustainable-investments-grow-25-23-trillion/> (First accessed: 24.01.19)
- Bognar, M. (2019) Binomial Distribution Calculator. University of Iowa. Available at: <https://homepage.divms.uiowa.edu/~mbognar/applets/bin.html>
- Brown, S.J., & Goetzmann, W. & Ibbotson, R.G. & Ross, S. A. (2015) Survivorship Bias in Performance Studies. *The Review of Financial Studies*. 5(4), p.553-580. Available at DOI: 10.1093/rfs/5.4.553
- Business Insider (2018) *Money managers are flocking to a \$23 trillion investing strategy that Morgan Stanley says is ready to take off* [Online]. Available at: <https://www.businessinsider.com/esg-investing-morgan-stanley-forecasts-strong-future-2018-6?r=US&IR=T&IR=T> [First accessed: 24.01.19]
- Carhart, M.M. & Carpenter, J.N, & Lynch, A.W. & Musto, D.K. (2002) Mutual Fund Survivorship, *The Review of Financial Studies*, 15(5) p.1439–1463. Available at DOI: 10.1093/rfs/15.5.1439
- Chatterjee, S. & Das, N. & Ruf, B. & Sunder, A. (2018). Fund Characteristics and Performance of Socially Responsible Mutual Funds: Do ESG Ratings Play a Role? *Journal of Accounting and Finance, Forthcoming*. Available at: arXiv:1806.09906
- Clark, G.L. & Feiner, A. & Viehs, M., (2015) *From the Stockholder to the Stakeholder: How Sustainability Can Drive Financial Outperformance* Available at SSRN: <https://ssrn.com/abstract=2508281>
- Dimson, E. & Marsh, P. & Staunton, M. (2015). Responsible investing: Does it pay to be bad? *Credit Suisse Global Investment Returns Yearbook* Available at: https://knowledge.insead.edu/sites/www.insead.edu/files/images/dimson_marsh_staunton_responsible_investing_1.pdf
- Dolvin, S.D. & Fulkerson, J.A. & Krukover, A. (2017). *Do 'Good Guys' Finish Last? The Relationship*

between Morningstar Sustainability Ratings and Mutual Fund Performance. Available at SSRN: <https://ssrn.com/abstract=3019403> or DOI: 10.2139/ssrn.3019403

Doyle, T.M. (2018) *Ratings that don't rate* [Online]. American Council for Capital Formation. Available at: http://accfcorgov.org/wp-content/uploads/2018/07/ACCF_RatingsESGReport.pdf (First accessed 10.02.18)

Eccles, R.G. & Verheyden, T. & Feiner, A. (2016). ESG for All? The Impact of ESG Screening on Return, Risk, and Diversification. *Journal of Applied Corporate Finance*, 28(2) p. 47-55. Available at SSRN: <https://ssrn.com/abstract=2808224> or DOI: 10.1111/jacf.12174

Elton, E.J. & Gruber, M.J. & Blake, C.R. (1996). Survivor Bias and Mutual Fund Performance, *The Review of Financial Studies*, 9(4) p.1097–1120. Available at DOI: 10.1093/rfs/9.4.1097

Eurosif – The European Sustainable Investment Forum (2019) *GSIA* [Online] Available at: <http://www.eurosif.org/our-network/gsia/> (First accessed: 30.01.19)

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), p. 383-417. Available at DOI: <https://doi.org/10.2307/2325486>

Fama, E. & French, K. (1993). Common Risk Factors in Returns on Stocks and Bonds. *Journal of Financial Economics*. 33. 3-56. Available at DOI: 10.1016/0304-405X(93)90023-5.

Fama, Eugene F. & French, Kenneth R., The Capital Asset Pricing Model: Theory and Evidence (2004). *Journal of Economic Perspectives* 18(3) p. 25-46. Available at SSRN: <https://ssrn.com/abstract=440920> or DOI: 10.2139/ssrn.440920

FDA – United States Food and Drug Administration (2018) *Family Smoking Prevention and Tobacco Control Act – An Overview* [Online]. Available at: <https://www.fda.gov/TobaccoProducts/GuidanceComplianceRegulatoryInformation/ucm246129.htm> (First accessed 04.02.2019)

Financial Times (2018) *Green finance: a contrarian take* [Online]. Available at: https://ftalphaville.ft.com/2018/11/15/1542294447000/Green-finance--a-contrarian-take/?fbclid=IwAR07GqNZVoSzyJnTW6LRGJrFgkAp-N-JywC7OTB1FDNFhc0BGxoK_vW_64 (First accessed 10.02.2019)

Forbes (2017) *Socially-Responsible Investing: Earn Better Returns From Good Companies* [Online]. Available at: <https://www.forbes.com/sites/moneyshow/2017/08/16/socially-responsible-investing-earn-better-returns-from-good-companies/#1c6c7210623d> (First accessed 03.04.19)

Fulton, M. & Kahn, B. & Sharples, C. (2012) *Sustainable Investing: Establishing Long-Term Value and Performance*. Available at SSRN: <https://ssrn.com/abstract=2222740>

Friede, G. & Busch, T. & Bassen, A., (2015), ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies. *Journal of Sustainable Finance & Investment*, 5(4). p.210-233. Available at DOI:10.1080/20430795.2015.1118916. or SSRN: <https://ssrn.com/abstract=2699610>

Giese, G. & Lee, L.-E. & Melas, D. & Nagy, Z. & Nishikawa, L. (2017) *Foundations of ESG Investing*. MSCI. Available at: <https://www.msci.com/documents/10199/03d6faef-2394-44e9-a119-4ca130909226> (First accessed 17.03.19)

Global Reporting Initiative (2019) *About GRI* [Internet] Available at: <https://www.globalreporting.org/information/about-gri/Pages/default.aspx> (First accessed 01.02.2019)

GSIA - Global Sustainable Investment Alliance (2016) *2016 Global Sustainable Investment Review* Available at: http://www.gsi-alliance.org/wp-content/uploads/2017/03/GSIR_Review2016.F.pdf (First accessed: 18.01.2019)

Han, J.-J. & Kim, H. J. (2016). Empirical Study on relationship between corporate social responsibility and financial performance in Korea. *Asian Journal of Sustainability and Social Responsibility*. Available at DOI: 10.1186/s41180-016-0002-3

Harvard (2017). *ESG Reports and Ratings: What They Are, Why They Matter*. Harvard University [Online]. Available at: <https://corpgov.law.harvard.edu/2017/07/27/esg-reports-and-ratings-what-they-are-why-they-matter/> (First accessed 31.01.2019)

Hill, R.P. & Ainscough, T. & Shank, T. & Manullang, D. (2007). Corporate social responsibility and socially responsible investing: A global perspective. *Journal of Business Ethics*, 70, 165-174. Available at DOI: [10.1007/s10551-006-9103-8](https://doi.org/10.1007/s10551-006-9103-8)

Hong, H. & Kacperczyk, M. (2007). *The Price of Sin: The Effects of Social Norms on Markets*. *Journal of Financial Economics*. Available at SSRN: 10.2139/ssrn.766465.

IIRC – The International Integrated Reporting Council (2018) *The IIRC* [Online] Available at: <http://integratedreporting.org/the-iirc-2/> (First accessed: 31.01.2018)

iShares (2019) *iShares ESG MSCI EM ETF*. [Online]. Available at: <https://www.ishares.com/us/products/283777/ishares-esg-msci-em-etf-fund>

Jensen, M.C. (1967), The Performance of Mutual Funds in the Period 1945-1964 *Journal of Finance*, 23(2) p. 389-416. Available at SSRN: <https://ssrn.com/abstract=244153>

Johnsen, T. and Gjølberg, O. (2007). *Etisk forvaltning av Staten Pensjonsfond Utland: En oppdatert analyse*. Available at: https://www.regjeringen.no/globalassets/upload/fin/vedlegg/aff/analyse_etisk_forvaltning.pdf

J.P. Morgan (2018) *Sustainable Investing is Moving Mainstream* Available at: <https://www.jpmorgan.com/global/research/esg> [First Accessed: 29.12.18]

John Hancock Investments (2019) *ESG Large Cap Core Fund* [Online] Available at: <https://www.jhinvestments.com/investments/esg/us-equity-funds/esg-large-cap-core-fund-c-jhjc.html> (First accessed 08.02)

Lesser, K., & Rößle, F. & Walkshäusl, C. (2016). International socially responsible funds: financial performance and managerial skills during crisis and non-crisis markets. *Journal of Problems and Perspectives in Management*, 14(3-2), 461-472. Available at DOI: 10.21511/ppm.14(3-2).2016.02

Kumar, N.C.A & Smith, C. & Badis, L. & Wang, N. & Ambrosy, P. & Tavares, R. (2016). ESG factors and risk-adjusted performance: a new quantitative model, *Journal of Sustainable Finance & Investment*, 6(4),p. 292-300, Available at DOI: [10.1080/20430795.2016.1234909](https://doi.org/10.1080/20430795.2016.1234909)

Lee, D.D. and Faff, R.W. (2009) Corporate Sustainability Performance and Idiosyncratic Risk: A Global Perspective. *The Financial Review*, 44(2), Available at SSRN: <https://ssrn.com/abstract=1344731>

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*. 7(1), p.77-91. Available at DOI: 10.1111/j.1540-6261.1952.tb01525.x

MSCI (2019). *ESG Ratings*. MSCI. Available at: <https://www.msci.com/esg-ratings> (First accessed 18.02.19)

Morningstar (2017) *The Morningstar Sustainability Rating. Helping Investors Evaluate the Sustainability of Portfolios* [Online] Available at: <https://www.morningstar.com/content/dam/marketing/shared/pdfs/sustainability/IWM17NovDec-MorningstarSustainabilityRating.pdf?cid=EMQ> (First accessed: 20.01.2019)

Morningstar (2018a) *Interpreting the Morningstar Sustainability Rating Changes* [Online]. Available at: <https://www.morningstar.com/articles/891279/interpreting-the-morningstar-sustainability-rating.html> (First accessed 22.01.2019)

Morningstar (2018b) *Morningstar Sustainability Rating Methodology* [Online]. Available at: https://www.morningstar.com/content/dam/marketing/shared/research/methodology/744156_Morningstar_Sustainability_Rating_for_Funds_Methodology.pdf (First accessed 23.01.2019)

Morningstar (2019) *About Us* [Online] Available at: <https://www.morningstar.com/company/about-us> (First accessed: 19.01.2019)

NBIM – Norges Bank Investment Management (2019) *Observasjon og utelukkelse av selskaper* <https://www.nbim.no/ansvarlighet/utelukkelse-av-selskaper/> (First accessed: 18.01.2019)

Nofsinger, J.R. & Sulaeman, J. & Varma, A. (2016). Institutional Investors and Socially Responsible Investments: It Just Makes (Economic) Sense. *Working paper*. University of Alaska. Available at: <http://capitalism.wfu.edu/wp-content/uploads/2016/04/Varma-paper.pdf>

Norges Bank (2019) *Statsobligasjoner Rente Årsgjennomsnitt av daglige notering*. Norges Bank. Available at: <https://www.norges-bank.no/Statistikk/Rentestatistikk/Statsobligasjoner-Rente-Arsgjennomsnitt-av-daglige-noteringer/>

Oikonomou, I. & Brooks, C. & Pavelin, S., (2010). The Impact of Corporate Social Performance on Financial Risk and Utility: A Longitudinal Analysis. *Financial Management* 41(2), p. 483–515. Available at SSRN: <https://ssrn.com/abstract=2172742>

Oslo Børs (2019) *Oslo Børs Mutual Fund Index*. Oslo Børs. Available at: https://www.oslobors.no/ob_eng/markedsaktivitet/#/details/OSEFX.OSE/overview

Regjeringen (2016). *Corporate Social Responsibility* [Online] Available at: <https://www.regjeringen.no/en/topics/foreign-affairs/business-cooperation-abroad/innsikt/csr/id2076260/> (First accessed: 01.02.2019)

Reuters (2017) *Investing with 'green' ratings? It's a gray area* [Online] Available at: <https://www.reuters.com/article/us-climate-ratings-analysis/investing-with-green-ratings-its-a-gray-area-idUSKBN19H0DM> (First accessed 29.01.2019)

Renneboog, L. & Horst, J. & Zhang, C. (2008). Socially Responsible Investments: Institutional Aspects, Performance, and Investor Behavior. *Journal of Banking & Finance*. Volume 32, Issue 9, p.1723-1742. DOI: 10.1016/j.jbankfin.2007.12.039.

Richardson, B.J. (2008) *Socially Responsible Investment Law: Regulating the Unseen Polluters*. Oxford University Press. Available at SSRN: <http://ssrn.com/abstract=1128704>

Sharpe, W. (1966). Mutual Fund Performance. *The Journal of Business*, 39(1), 119-138. Available at: <http://www.jstor.org/stable/2351741>

Shiller, R. (2003). From Efficient Market Theory to Behavioral Finance. *Journal of Economic Perspectives*. 17(1). 83-104. Available at DOI: 10.1257/089533003321164967.

Sustainalytics (2017) *About Us* [Online] Available at: <https://www.sustainalytics.com/about-us/#History> (First accessed: 19.01.2019)

Sustainalytics (2018) *Controversies Research* [Online] Available at: <https://www.sustainalytics.com/controversies-research/> (First accessed: 20.01.19)

Swedroe, L. (2018). *Explaining the Beta Anomaly* [Online] Alpha Architect Available at: <https://alphaarchitect.com/2018/06/28/explaining-the-beta-anomaly/> (First accessed 09.04.19)

United Nations Global Compact (2004) *Who Cares wins. Connecting Financial Markets to a Changing World*. Available at: https://www.unglobalcompact.org/docs/issues_doc/Financial_markets/who_cares_who_wins.pdf (First accessed: 27.12.2018)

United Nations Principles for Responsible Investment (2019) *About the PRI* [Online] Available at: <https://www.unpri.org/about-the-pri> (First accessed: 21.01.2018)

US SIF - The Forum for Sustainable and Responsible Investment (2019) *SRI Basics* [Online] Available at: <https://www.ussif.org/sribasics> (First accessed 24.01.19)

US SIF - The Forum for Sustainable and Responsible Investment (2018) *Sustainable investing assets reach \$12 trillion as reported by the US SIF Foundation's biennial Report on US Sustainable, Responsible and Impact Investing Trends*. Available at: <https://www.ussif.org/files/US%20SIF%20Trends%20Report%202018%20Release.pdf> (First accessed: 30.01.2019)

Utz, S., (2017) Corporate Scandals and the Reliability of ESG Assessments: Evidence from an International Sample. *Review of Managerial Science, Forthcoming*. Available at SSRN: <https://ssrn.com/abstract=3108306>

Vanguard (2019) *Vanguard ESG U.S Stock ETF*. Vanguard. Available at: <https://investor.vanguard.com/etf/profile/ESGV> (First accessed 08.02.19)

Varma, A. & Nofsinger, J.R., (2012). Socially Responsible Funds and Market Crises. *Journal of Banking and Finance, Forthcoming*. Available at SSRN: <https://ssrn.com/abstract=2142343>

Washington Post (2012) *Everything you need to know about the assault weapons ban, in one post* [Internet] Available at: https://www.washingtonpost.com/news/wonk/wp/2012/12/17/everything-you-need-to-know-about-banning-assault-weapons-in-one-post/?utm_term=.51854c41fea6 (First accessed 04.02.2019)

Wimmer, M. (2012). ESG-persistence in Socially Responsible Mutual Funds. *Journal of Management and Sustainability*, 3(1), p.9-15. Available at DOI: <http://dx.doi.org/10.5539/jms.v3n1p9>

10. Appendix

Table A1. Portfolio allocation by fund category.

	Category	Europe Eq. Mid/Small	Global Eq. Large Cap	Europe Eq. Large Cap	Aggressive Allocation	Sector Equity	Global EM Equity	Equity Misc.	Other Equity(*)
HPSS	High	12	1	16	0	0	0	0	0
	Low	2	9	0	1	3	7	2	5(3)
Contr	High	25	1	0	0	1	0	2	0
	Low	0	15	4	0	3	0	4	3
Environ.	High	4	3	18	0	4	0	0	0
	Low	7	6	0	2	3	7	2	2
Social	High	8	0	21	0	0	0	0	0
	Low	3	6	0	0	4	7	2	7(3)
Govern.	High	12	0	17	0	0	0	0	0
	Low	2	8	0	0	4	7	2	6(3)

Table A1 portrays the allocation of fund categories in the top and bottom quintile portfolios.

(*) The number in parenthesis covers the number of funds categorized "US Equity Large-Cap Blend", of which there are 3 funds total in the 146 funds sample. These funds are included in the statistics for total Large-Cap funds presented in section 6.5.

Table A2. Descriptive statistics of sustainability metrics

	Sustainability score			Controversy Score		
	Full sample	High	Low	Full sample	High	Low
Mean	51,83	57,95	46,00	4,43	1,75	7,84
Median	50,86	57,94	46,19	4,14	1,94	7,65
Std.dev.	4,32	0,85	1,36	2,24	0,42	0,94
Max.	59,70	59,70	47,98	10,54	2,19	10,54
Min.	43,13	56,37	43,13	0,89	0,89	6,93

	Environment			Social			Governance		
	Full sample	High	Low	Full sample	High	Low	Full sample	High	Low
Mean	55,95	59,75	51,53	56,03	61,00	51,00	56,04	62,10	49,86
Median	56,41	59,78	51,77	56,00	61,55	51,40	55,41	61,98	50,40
Std.dev.	2,85	1,03	1,57	3,51	1,38	1,71	4,44	1,07	1,62
Max.	61,40	61,40	54,01	63,78	63,78	53,21	64,97	64,97	51,95
Min.	45,85	58,00	45,85	47,80	58,90	47,80	46,01	60,41	46,01

Table A2 displays the descriptive statistics for a variety of sustainability metrics. The full sample consists of 146 funds, while the High and Low columns contain statistics for the top and bottom 29 funds (20%) respectively.

Table A3. Cross-sectional regression results: Sub-sample, average returns with respect to beta and HPSS.

<i>Regression statistics</i>			
Adj. R-square		0,15	
Standard error		0,02	
Observations		46	
	<i>Coefficients</i>	<i>Standard error</i>	<i>t-Stat</i>
Intersection	0,24***	5,40 %	4,49
Beta	0,003	1,95 %	0,13
HPSS	-0,003***	0,10 %	-2,71

Table 7.5 shows the regression results for a cross-sectional analysis after isolating a portion of the sample that closer reflects the selected benchmark. N=46 funds with average annual returns as LHS variables and beta and HPSS as RHS variables. Stars signify a coefficient value statistically significant as follows: () = 10% level, (**) = 5% level, (***) = 1% level.*

Table A4. Cross-sectional regression results: Average returns with respect to beta, HPSS, geographical dummy and top and bottom sustainability dummies.

<i>Regression statistics</i>			
Adj. R-square		0,28	
Standard error		0,03	
Observations		146	
	<i>Coefficients</i>	<i>Standard error</i>	<i>t-Stat</i>
Intersection	0,35***	0,101	3,44
Beta	-0,106***	0,014	-7,51
HPSS	-0,000**	0,002	-2,03
Europe	0,041***	0,011	3,67
TopDummy	0,010	0,012	0,88
BotDummy	-0,026**	0,012	-2,20

Table A4 shows the regression results for a cross-sectional analysis after adding a variety of dummy variables to the 7.4 regression. Stars signify a coefficient value statistically significant as follows: () = 10% level, (**) = 5% level, (***) = 1% level.*

Table A5. Cross-sectional regression results: Average returns with respect to betas and ESG-momentum.

<i>Regression statistics</i>			
Adj. R-square	0,24		
Standard error	0,04		
Observations	146		
	<i>Coefficients</i>	<i>Standard error</i>	<i>t-Stat</i>
Intersection	0,135***	0,009	14,65
Beta	-0,081***	0,012	-6,60
ESGmomentum	0,011***	0,004	2,64

Table A5 shows the regression results for the ESG-momentum factor. Stars signify a coefficient value statistically significant as follows: () = 10% level, (**) = 5% level, (***) = 1% level.*



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway