Erasmus University Rotterdam

Erasmus School of Economics

Master Specialization Financial Economics

MASTER THESIS

Socially Responsible Investing and Portfolio Performance

Author: **Marketa Pokorna**

Student number: **449106**

Supervisor: **Dr. Esad Smajlbegovic**

Second assessor: **Dr. Maurizio Montone**

Finish Date: **November 9, 2017**

Abstract

This thesis contributes to investigating the performance of investment strategies falling under the area of Responsible Investing. The results of this study are important for the investment decision-making of both values-driven and profit-driven investors. The empirical analysis examines returns of trading strategies based on company corporate social responsibility as measured by Environmental, Social and Governance indicators (ESG) obtained from the Thomson Reuters ESG database. Strategies excluding sin stocks and holding long position in top ESG-rated stocks and short position in bottom-rated ESG stocks represent the strategy of the environmentally and socially conscious investors. Such strategy harms the investor's portfolio performance and a reverse strategy of holding long position in bottom ESG stocks and short position in top ESG stocks leads to positive abnormal returns of up to 6% per year. These abnormal returns cannot be attributed to a sentiment-driven mispricing. Instead, they are likely to be a compensation for low-sustainability risk. The results suggest that the values-driven investors should not restrict their investment choices into public equities by considering the ESG ranking. They should rather fully diversify their portfolios and then use the proceeds to invest into projects that comply with their personal values and convictions.

JEL Classification	G11, G12, G41, M14
Keywords	Socially responsible investing, ESG, Investor sentiment

Author's e-mail	marketa.pokorna.jh@gmail.com

Contents

List of Tables

List of Figures

Acronyms

RI Responsible Investing

SRI Socially Responsible Investing

CSR Corporate Social Responsibility

GIIN GLobal Impact Investing Network

PRI Principles for Responsible Investment

ESG Environmental, Social, Governance

NGO Non-Governmental Organization

ISIN International Securities Identification Number

KLD Kinder, Lyndenberg, and Domini database

OTC Over-the-counter

TRBC Thomson Reuters Business Classification

BW Baker and Wurgler sentiment index

CAPM Capital Asset Pricing Model

SMB Small Minus Big

HML High Minus Low

WML Winners Minus Losers

CMA Conservative Minus Aggressive

Chapter 1

Introduction

Investors have traditionally pursued financial return as a measure of the outcome of their investments. However, they have recently also started examining activities of companies they invest in and the positive or harmful effects that these companies may have on the society and environment. This approach is a new concept which contradicts the classical risk-return relationship in investment decision making where the expected return is maximized for a certain level of risk or, alternatively, the risk is minimized for a desired level of expected return.

The alternative view on investments stems from the changing world set-up. It was initiated by the more pressing threat of global warming and a number of environmental disasters in the recent past. International treaties such as the Kyoto protocol or Paris agreement have helped to stimulate the change of thinking. Also, persisting poverty issues and widening wealth gap have emphasized the importance of investment into social capital.

Not only changing environment and conditions but also changing generation is significant for the future of such alternative investments. The generation of so called 'millennials' is likely to be important for investment industry in the near future. These young individuals are usually characterized as liberal, socially conscious and promoting their values to shape the world since they have the majority of their lives lying ahead. Moreover, millennials are expected to inherit considerable wealth and so acquire more funds to be invested within the next two decades. In the United States, about $40 trillion of wealth is supposed to be transferred to more than 75 million millennials within the next 20 years (Chasan, 2017). All these trends indicate that the future nature of investments might not be purely profit seeking.

The concept of Socially Responsible Investing (SRI) is perceived as one of the solutions to tackle this issue. This investing approach takes into account the side-effects of directing funds to certain companies. It is based on screening the potential companies according to their corporate social responsibility (CSR). Such screening can be done by assessing the company's behavior towards dealing with environmental, social and corporate governance (ESG) issues. Values-driven investors are able to use the ESG criteria to assess the companies according to what they perceive as company responsibility and sustainability. Then the question arises whether such investments are merely for environmentally and ethically conscious investors who are likely to be willing to give up part of their financial returns to follow their values.

The purpose of this thesis is to examine the financial performance of investment strategies based on company corporate social responsibility. It is desirable to provide evidence whether investments in stocks of companies with decent corporate social responsibility bring significantly different returns from a benchmark represented by a portfolio that is diversified. Answering this question is crucial as it indicates if the CSR-minded investments are suitable only for people who strongly value the impact of their investments. Alternatively, these investments can be attractive also for purely profit-seeking investors. It can also be argued that a value-minded investor would not invest in companies connected with harmful industries. For this reason it will also be investigated if removing so called 'sin stocks' from the portfolio affects the performance.

This thesis makes use of the ESG database compiled by Thomson Reuters.Simple trading strategies are designed based on signals from ESG and then their performance is evaluated. The strategy of holding long position in good ESG stocks short and position in low ESG stocks results in returns significantly lower than zero. The results are the same in the strategy when the sin stocks are removed from he portfolio. These findings do not change when the returns are properly adjusted for risk and suggest that to reach positive abnormal returns, the investor would have to buy the low ESG stocks and short the high ESG stocks. The results suggest that the abnormal returns cannot be attributed to a sentiment-driven mispricing. Instead, the abnormal returns are viewed as a compensation for low-sustainability risk.

The results offer a view that could curb the value-minded investors' engagement in Socially Responsible Investing. By investing in companies with good corporate social responsibility and financing this from a short position in bad sustainability companies, the investor's portfolio performance would experience a severe harm. Therefore, these investors are advised to follow their personal values by diversifying their portfolio and then invest the proceeds in projects with positive impact that is in line with the investor's beliefs. On the other hand, purely profit-seeking investors could make use of the active management in stock picking by utilizing the ESG score to construct long-short strategies that buy the low ESG stocks and short the high ESG stocks.

Chapter 2

Literature and Background Review

This chapter gives all necessary concept definitions as well as industry overview and recent development of investment industry concerning the focus of this thesis. After that all possible relationships between the CSR and company financial performance are discussed together with their reasoning. Then this chapter reviews the relevant academic research in the area of this thesis topic and ends with the hypothesis development.

2.1 Concept development and definition

Even though socially responsible investments are a relatively new field, the interest for this area is rising rapidly. However, the lack of professionals being specialized in this industry as well as the low regulatory and legal framework make it common to use several terms interchangeably. Generally speaking, investing where not only profit maximization is pursued, represents investments that are to some extent in line with investor's personal values. Among the terms describing the same general idea are social investing, green investments, impact investing, ethical investments, aware and conscious investing, responsible investing, or sustainable investments. Vague boundaries among these concepts and the absence of clear definitions has been discussed among the researchers (Höchstädter and Scheck, 2015). However, clearer picture on the terms is arising as agencies involved in these areas are becoming more active and more researchers are attracted to these topics.

All the terms stated above describe investments into companies where the investor tends to consider company's corporate social responsibility (CSR) or the impact resulting from its business activities. The latter one is the recently increasingly popular field of impact investing. The Global Impact Investing Network (GIIN) is a non-profit organization whose goal is to improve effectiveness of impact investing globally. It defines impact investing as investments made into companies for the purpose of intentionally generating social or environmental impact as well as a financial return. An example of an impact investment could be investing in a company which encourages underprivileged students to study in college. If the college pays the company for student support trainings, then there is a financial return as well as social return. Hebb (2013) states that impact investing brings an innovative way how to deal with some crucial world issues and at the same time provides the desired financial returns demanded by investors. However, the author also argues that impact investing is mostly a micro-finance field so far and for its expansion, it is necessary to establish a coordinated marketplace.

Another problem preventing research from analyzing the financial performance of impact investments is the difficulty of measuring and quantifying the non-financial impact and the value created. There is a progress in this issue made by GIIN in terms of compiling a catalogue of generally accepted performance metrics on social and environmental performance of companies. Several studies have been published so far to examine the impact investing industry, such as by interviewing key actors in Mendell and Barbosa (2013). However, no attempt has been made yet to measure the impact of investments in the empirical research. As Höchstädter and Scheck (2015) conclude, important issues concerning the identification of impact companies and measuring the resulting impact have to be solved so that this investing area reaches a higher level of credibility.

Next to the impact investing we can find another branch of the responsible realm, which is socially responsible investing (SRI). Unlike impact investing, where positive impact from the investment is intended, SRI abstains from seeking impact created *as a result* from business activities and rather pursues responsible behavior of companies *during* and *alongside* their business activities. This concept, where companies take responsibility for their effect on the environment and society, is known as company's corporate social responsibility (CSR) and therefore the term CSR commonly accompanies the term SRI in the literature. The last important term connected to SRI and CSR is the concept of measuring CSR. It is usually referred to as environmental, social, and governance (ESG) characteristics of a company.

The difficulty of assessing the resulting impact in impact investing represents a barrier to directing investor's funds to the higher impact projects. In contrast to that, SRI can channel one's investments to companies with better CSR, measured by ESG, and therefore provides the opportunity for empirical research of these investments.

The leading proponent of socially responsible investing is nowadays the United Nations' partnership association called PRI (Principles for Responsible Investment). This investor initiative sets the global view on responsible investing and the majority of academics follow PRI's terminology. PRI defines Responsible Investing as "an approach to investing that aims to incorporate ESG factors into investment decisions, to better manage risk and generate sustainable, long-term returns". Notably, responsible investing does not require ethical considerations. It simply incorporates ESG factors because not doing so would imply ignoring risks that might influence the returns. Therefore, responsible investing represents a concept broader than SRI as it can be a strategy of an investor who is focused merely on the financial return. According to PRI, the socially responsible investment is a subset of responsible investing in which financial return is combined with moral or ethical considerations. Throughout this thesis, the terms will be used in line with the PRI framework. Nevertheless, when doing an empirical research on investment styles, investor's values do not play role and therefore the difference between SRI and responsible investing can be put aside until the interpretation of the results.

The PRI was launched in April 2006 by 100 founding signatories. Since then the number of signatories has grown to more than 1600. The signatories join the initiative voluntarily and commit themselves to comply with six investment principles. Based on them, signatories are obliged to incorporate ESG into their investment decisions and report their activities of implementing the principles. The aim of creating the network of signatories is to help develop a sustainable global financial system. The progress of this project is not only expressed by the growth of participants but also by the assets under management. Since 2006, signatories' assets under management have increased from $6.5 trillion to $62 trillion in 2016, where half of this growth happened during the last 3 years (PRI, 2017).

The rise of responsible investments is also documented by the attention paid to several newly constructed market indices. Among the ones most commonly watched are S&P ESG Indices, DJ Sustainability World Composite or MSCI World ESG Index. Another example of the increasing importance of ESG characteristics is the emergence of loans where interest rate is tied with the sustainability performance as measured by ESG criteria. The pioneer in this area is ING Groep NV who designed a unique loan repayment structure based on the borrower's ESG score achievement. The loan's interest rate is supposed to change every year according to the ESG ranking of the borrower (Pardini, 2017). This might represent a desirable solution because it decreases risk for lenders, sets better conditions for firms with good ESG performance, and supports responsible behavior of companies and therefore adds value to the society. The last current example of the growing ESG importance is ESG integration by the Swiss Re Group. Swiss Re integrates ESG into its investment processes since the start of 2017. However, in July 2017 they announced they switched to ESG benchmarks for their portfolios worth $130 billion. The company is the first one in the re/insurance industry to shift their benchmarks to the alternative ESG ones. By such a step it suggests that considering ESG is not only an addition to its processes but it is now an integral part (Revill, 2017).

As described above, responsible investments are getting more and more attention nowadays. However, the majority of these investments is made by institutional investors such as pension funds and other institutions. Only 26 percent of SRI was attributable to individual investors at the beginning of 2016, up from 13 percent in 2013 (Colby, 2017). The involvement of institutions in responsible investing could be partly explained by building up reputation and stakeholders' relations. For the retail investors, ethical values probably play the role. For more investors to become attracted, the financial performance of SRI needs to be examined. The next section will review some theories on how investing with CSR in mind could affect the investor's financial performance.

2.2 Existing hypotheses about SRI returns

Studying the relationship between financial and social/environmental performance of an investment strategy is interesting for two reasons. Firstly, the investors who invest solely based on their values are sometimes willing to sacrifice financial return to adhere to their values. With a positive relationship between CSR and investment performance, they could also reap financial returns alongside their moral satisfaction. Secondly, many investors think about company's ethics only as a random side effect and are still mainly interested in the financial return. Then, even these investors who do not have non-financial interests could make use of the potential relationship, direct their funds more effectively and reach the desirable level of returns.

One of the main arguments against socially responsible investments is based on Milton Friedman's view of the social responsibility of business (Friedman, 1970). This critique states that if it is not possible for investors to achieve desirable financial return with SRI (direct way), they could still follow their ethical values by investing into diversified portfolio and then using part of the financial returns to invest in projects which represent their values (indirect way). For example, instead of giving less weight to (or excluding) a company that violates employee gender equality, an investor could diversify and use the returns to invest in projects promoting women' employment. Therefore, whether strategies based on company's good CSR affect financial performance of portfolios positively, neutrally, or negatively is important for the investor to choose between the direct and indirect option.

The first relationship that could be expected is a negative impact of high CSR on portfolio performance. On a company level, higher CSR might imply competitive disadvantage because it represents costs that could be avoided. These costs reduce profits and therefore also the wealth of shareholders. This concept is often referred to as the agency view of corporate social responsibility, i.e. when pursuing CSR, the interests of managers and shareholders are diverging. Under this view, CSR is detrimental for shareholders but is pursued by managers, who fail to obtain compensation for good firm performance, in sight of private benefits (awards and other appreciations from promoters of social responsibility). This expectation therefore views CSR as a waste of corporate resources (Ferrell et al., 2016). When considering a portfolio which invests in companies with good CSR (i.e. excludes certain companies that normative portfolio theory would include), it can be argued that under-diversification would lead to sub-par returns.

The second possible association is a neutral one. The expectation here is that the company's CSR activities increase the costs and benefits by a similar amount. This relationship is supported by Ullmann (1985) who describes that so many factors influence the relationship between CSR and financial performance that no prevailing effect can be expected. Also, any resulting relationship could simply be a coincidence. However, Ullmann (1985) states that neutral relationship could also result from the lack of empirical data concerning this topic. This can consequently disguise any relationship which is there.

A positive link is the third plausible relationship. One argument could be that good CSR results from exceptional management skills which also lower the costs. Alternatively, high level of CSR reduces future risks of corporate scandals or lawsuits concerning negative externalities of the company. That leads to higher expected returns in the future. Another approach is adopted by Derwall et al. (2011). They develop a hypothesis called the errors-in-expectations hypothesis. It argues that CSR provides an information about the company's intrinsic value, which is not fully understood by the financial markets. This then generates abnormal returns until all relevant information is reflected in the stock prices.

The last relationship that is hypothesized is the link between socially controversial stocks and financial performance. These stocks are usually referred to as 'sin stocks', 'controversial stocks', or 'vice stocks'. Such companies are not recognized by CSR but rather by operating in industries which are generally considered controversial. Derwall et al. (2011) formulate a so called shunned-stock hypothesis which assumes that investors following certain values exclude these stocks and therefore create a shortage of demand for these stocks. This effect can be explained by the model of incomplete information and segmented capital markets of Merton (1987). The segmented markets arise due to information asymmetry. Certain stocks are therefore ignored by investors and are traded with a discount because fewer investors follow them. As Hong and Kacperczyk (2009) state, these stocks are neglected especially by institutional investors who are usually obliged to follow strict rules to choose investable industries. The prices of such stocks then get lower compared to their fundamental values. Therefore, this hypothesis predicts that sin stocks would have higher expected returns. Moreover, companies with potentially harmful products face higher litigation risks which could be reflected in the expected returns (Fabozzi et al., 2008).

2.3 Performance of socially (ir)responsible portfolios

In this section the academic literature on financial performance of responsible investments will be reviewed. Even though the first attempts made to discover the relationship were already in 1980s and 1990s, the construction of new models (e.g. Fama and French (1993) and Carhart (1997)) and the development of better data collection techniques facilitated more sophisticated modelling to be used later on.

One of the most widely used approaches is to invest in companies with good CSR, measured by ESG characteristics. Here the long position is held in the companies with the best ESG ranking and the short position is held in the companies with the worst ESG ranking. This is also often termed as positive screening. Kempf and Osthoff (2007) provide evidence that portfolio of companies with high ESG attributes demonstrates abnormal returns during sample period 1992 to 2004 and Statman and Glushkov (2009) find the same for sample period from 1992 to 2007. These results are in line with the errors-in-expectations hypothesis. However, Halbritter and Dorfleitner (2015) perform an extensive study comparing different data sources and concludes that the abnormal returns produced by this strategy are statistically insignificant. This result is robust to applying different weighting structures (equal, value) and ESG thresholds for the portfolio. Auer and Schuhmacher (2016) use a new database which re-evaluates the ESG scores more frequently and the findings are in line with Halbritter and Dorfleitner (2015), i.e. the active choice of high ESG stocks does not bring superior risk-adjusted returns.

Some studies have focused on national markets or portfolios formed only by environmental or social characteristic. For example Guenster et al. (2011) focuses on portfolio formed by environmental leaders and finds that investors do not need to choose between environmental and financial performance as the information on companies' eco-efficiency is exploitable to achieve higher returns. Interestingly, Brammer et al. (2006), who focuses on the market in the United Kingdom, provides evidence on a negative relationship, which suggests that CSR investing might be detrimental to shareholders. However, they examine only a short time period from 2002 to 2006.

Another commonly used method in selecting stocks is investing in high ESG stocks in combination with a so called negative screening. As in the first approach, high ESG companies are selected but they are screened for industry specifics and companies operating in industries perceived as irresponsible are excluded from the sample. This is a stricter form of SRI as it uses twofold ethical considerations (industry of operation and ESG). Kempf and Osthoff (2007) propose a simple trading strategy of buying high ESG rating and shorting the stocks with low ESG ratings. The results suggest that one can earn high abnormal returns with this strategy and that the approach works the best in combination with the above mentioned negative screening.

Contradicting findings are provided by Statman and Glushkov (2009). They state that the benefits of investing in high CSR scores are lost when certain stocks are excluded by negative screening. Therefore, they give the recommendation to refrain from excluding stocks from the high ESG portfolio. Similarly, Adler and Kritzman (2008) perform simulations to examine the costs incurred by excluding companies. They found that negative screening is accompanied by substantial costs.

Another approach is to perform only negative screening, without any ESG considerations. That simply means that irresponsible industries are excluded. Such portfolio does not comply with ethical investor's values, it is merely a weak form of responsible investment. These portfolios are often referred to as 'no harm' portfolios. Alternatively, one can investigate the financial performance of strategies based on investing into companies from industries that are perceived as irresponsible. Such attitude tests the shunned-stock hypothesis and is often called investing in sin stocks, because it selects businesses involved in activities usually considered as sin-seeking. These mainly include industries such as alcohol, tobacco, oil, or gaming.

There is a considerable academic interest in sin stock portfolio performance. Sin industries are usually connected with moral controversy or environmental, social, or ethical problems. Academic research on sin stocks is particularly interesting because there is no consensus on which industries should be classified as sin. Usually, studies consider companies involved in activities connected with alcohol, tobacco, gambling, oil, weapons, biotechnology, sex industry and nuclear. However, these differ among studies depending on the author's definition of 'sin' and availability of data. Recently, Blitz and Fabozzi (2017) mention also a new class of sin - marijuana sin stocks. They also state that more activities could be classified as 'sin' in the near future, e.g. for-profit prisons, predatory lenders and companies using sweatshops (violating labor laws).

Fabozzi et al. (2008) examine the performance of sin stock investing for 267 sin stocks in 21 national markets. They report that the sin portfolio significantly outperforms common benchmarks and thus the results are in conformity with the shunned-stock hypothesis. A similar study is carried out by Hong and Kacperczyk (2009) who also show that sin stocks exhibit abnormal returns. However, they attribute this not only to the idea that sin stocks are neglected but also to the higher litigation risk faced by these companies.

The superior returns of sin stocks are also found by Kim and Venkatachalam (2011). They note that sin companies have better earnings quality compared to control groups, because they are subject to a more intense regulatory scrutiny. A ground-breaking finding has recently come from Blitz and Fabozzi (2017) who re-examine the sin stock portfolio performance and conclude that there is no abnormal return after proper risk adjustment. Their analysis is in line with the previous findings considering classical (risk) factors: size, value and momentum[1]. However, the abnormal returns vanish when they control also for the Fama and French (2015) quality factors: exposure to profitability and investment.

An innovative screening approach is chosen by Cai et al. (2012). They consider investing in sin companies with good CSR. This is by no means contradicting; sin companies with good CSR exist as companies might aspire to improve their image by good CSR. Even though the study does not examine the financial performance of portfolios, they indicate that CSR behavior of sin companies increases the firm value. There is, however, no evidence yet on financial performance of portfolios with this screening procedure. A related research was conducted by Kim et al. (2014) who consider a universe of stocks with data on CSR (not only sin stocks), and find that stocks with better CSR do not behave responsibly to cover up bad news. Instead, future stock price crash risk is reduced with good CSR. However, this might not be the case when a portfolio of sin stocks is considered because such companies might have bigger incentive to divert the stakeholders' scrutiny.

2.4 Hypothesis development

To summarize the literature review, Table 2.1 presents an overview of criteria considered when following different investment approaches.

The table describes different styles of investment by the factors that investors take into consideration. These are whether an investor makes industry screening, integrates ESG characteristics, makes the investment with ethical values in mind and whether the investor considers the impact resulting from the investment. Both responsible investing and SRI integrate ESG, but SRI is also made with ethical values in mind. It is important to note that these two terms carry the different terminology because of the investor's ethical intention, however, cannot be distinguished when making a technical analysis of the portfolio performance. As a result, the empirical analysis in this thesis will bring

[1]When speaking of the factors from the model of Carhart (1997), it would be misleading to refer to them collectively as risk factors. This is due to the fact that the momentum factor's link to the true risk is unclear, however, the factor seems to be useful in explaining returns and is therefore added to the model.

Table 2.1: Classification of investments

	Industry screening	ESG integration	Investor's values	Resulting impact
Responsible investing	no	yes	no	no
SRI	no	yes	yes	no
SRI strict form	yes	yes	yes	no
Impact investing	no	no	yes	yes
No harm	yes	no	yes	no
Sin stocks	yes	no	no	no
Responsible sin	yes	yes	no	no

results that apply to both responsible investing (i.e. ESG integration for financial return) and SRI. Next, the strict form of SRI conducts the negative screening on companies and then incorporates ESG. Impact investing does not integrate ESG, it simply seeks resulting impact. No harm investing screens industries and excludes the stocks form the controversial industries. In contrast to that, sin stock investing is performed to achieve high expected returns by screening industries and choosing sin stocks. Lastly, 'responsible sin' investing follows sin industries and invests in those stocks based on ESG characteristics.

To review the findings from the literature, no clear answer has been given on the performance of SRI by examining ESG-based investment styles. There is no unanimous evidence whether ESG integration is able to produce abnormal returns, especially because several studies find the abnormal returns to be insignificant. Even more confusion is about the ESG integration combined with negative screening, where evidence on both positive and negative effect on the portfolio performance can be found in the literature.

Overall, there is a need for further studies to replicate these results to examine the sign and significance of the relationship. Moreover, it is necessary to review the relationship by using a more recent dataset. This urge follows from the works of Derwall et al. (2011) and Bebchuk et al. (2013) who both claim that the abnormal returns of CSR have been declining over time. Also, majority of studies are using data samples only until 2007. However, there has been an immense development of the responsible investments in the last decade and therefore the current sample is the most reflective.

Apart from re-assessing the CSR-financial performance relationship with a current dataset, there are other gaps in the literature. Already the work of Rowley and Berman (2000) points out that the crucial question for this relationship is not only *whether* to invest in high CSR companies but *when* or under what conditions. It is reasonable to suppose that market sentiment influences how strongly investors stick to their values and whether they follow certain stocks during low or high sentiment periods. Therefore, this offers the possibility to examine whether the potential returns can be explained by the level of a market sentiment index. In this way, it can be seen whether the sentiment could be a partial explanation for the abnormal returns that are left after accounting for the traditional risk factors of Fama and French (1993).

This expectation is based on the research of Stambaugh et al. (2012) who find that several anomalies are stronger in months following periods of high sentiment. It stems from the fact that overpricing should be more pronounced than underpricing since the noise-traders drive the price up, stock becomes overvalued and short-sale constraints prevent the rational traders from setting the price back. In case that the sentiment can be related to the style returns, this would indicate the presence of a sentiment-driven mispricing. Excess returns then would not reflect exposure to any additional risk factors apart for the risk factors already included in the model. In case that the sentiment cannot be related to the style returns, this result would suggest that there might be mispricing of a different form or other macroeconomic risks behind the excess returns. Contrary to Stambaugh et al. (2012), Sibley et al. (2016) claims that the relation between sentiment and the anomaly returns is driven by the component of the sentiment which is connected with the business cycle. Therefore, the sentiment has to be separated from the business cycle effects so that it can be used for the analysis.

The following chapter will describe the research design set up to answer the main question of this thesis: Is socially responsible investing attractive only for the environmentally/socially conscious investors or could it also attract purely profit-driven investors? First part will focus on examining the portfolio performance for an investment strategy utilizing ESG information on the stocks of companies traded. Second part will test whether incorporating the market sentiment helps to explain the result from the first part of the analysis. This could indicate whether the potential abnormal returns of ESG based investment strategies arise as a compensation for risk connected with ESG or whether they could be attributed to a sentiment-driven mispricing.

Chapter 3

Research Design

This chapter describes the data and methodology used for this thesis. All research in responsible investing is subject to existing databases on ESG data. Therefore, an overview of different databases will be given together with the description of the data source used for this thesis. Next, the methodology part will describe different investment strategies and portfolio construction. Finally, it gives a model for testing the hypotheses and outlines some robustness tests.

3.1 Data

Reporting information on ESG characteristics has been a voluntary initiative of companies since its beginning. Until today there is no legal obligation for companies to disclose data for a broad scale of ESG criteria. Therefore, databases to be used for responsible investing analysis do not exist for the whole universe of stocks. However, several rating agencies contributed to development of decent ESG data sources. The agencies specialize in gathering and evaluating ESG data from several sources such as company websites, annual reports, NGO reports, CSR reports, stock exchange filings, or the media. They collect information on various categories which belong to CSR. In the end, one final score is produced to give a cohesive ESG score for a particular company.

The longest database of ESG is recorded since 1991 by MSCI ESG metrics (formerly known as Kinder, Lyndenberg, and Domini Research and Analytics Inc. (KLD) database). This database gives information on strengths and weaknesses of companies in different categories, recorded as binary information (e.g. $strength_i = 1$, $weakness_j = 0$). The total score is then a simple sum of strengths less the sum of weaknesses of a company. However, Mattingly and Berman (2006) examine the methodology of this approach and find the simple aggregation of strengths and weaknesses insufficient and suggest to reconsider previous studies using this method.

Another widely used database is the ASSET4 ratings provided by Thomson Reuters. With history from 2002, it provides nowadays data on 280+ key performance indicators concerning ESG for 5000+ companies. In this database, the binary response for a certain indicator is translated to a percentage by a z-scoring procedure. A standard score (z-score) expresses the value in units of standard deviation form the mean value of all companies and produces a final percentage score for a certain company. Four pillars (environmental, social, corporate governance and economic) are equally weighted throughout this method. Apart from MSCI ESG and ASSET4, other sources are sometimes used, including databases compiled by Sustainalytics, Bloomberg L.P., RobecoSAM, or Ethical Investment Research Services (EIRIS).

In this thesis an emerging dataset by Thomson Reuters will be used. The Thomson Reuters ESG scores for public companies is a dataset which differs from the previously widely used ASSET4 database. Compared to the Thomson Reuter's ASSET4 database, this dataset is an enhancement and uses an improved methodology (Thomson Reuters, 2017). The ASSET4 data is accessible from Thomson Reuters Datastream while the Thomson Reuters ESG improved database is only available from the Thomson Reuters' EIKON software (successor of the Datastream). To the best of my knowledge, this dataset has not been used yet in the academic research to evaluate trading strategies based on ESG criteria.

The Thomson Reuters ESG database contains 6000+ global public companies with annual ESG information going back to 2002. Particular stock indexes were being included gradually. At the beginning, the database started with the SMI, DAX, CAC 40, FTSE 100, FTSE 250, S&P 500, and NASDAQ 100. Then other indexes were added: DJ STOXX, MSCI World (2008), S&P/TSX COMPOSITE (2009), Russell 1000, MSCI Emerging Markets (2011), Bovespa (2012), S&P ASX 300 (2013), and S&P NZX 50 (2016). Recently, the companies of Russel 2000 index are being included.

The dataset covers 400+ metrics which come from company public disclosures. The information is aggregated into 10 topics and then to 3 categories: Environmental category (Resource Use, Emissions, Innovation), Governance category (Management, Shareholders, CSR strategy) and Social category (Workforce, Human Rights, Community, Product Responsibility). The description of the categories can be found in Table B.1 in the Appendix B. The overall ESG score is generated from these three categories, it ranges between 0 and 1 and the higher the score, the better the ESG performance.

Apart from this score, the database offers the so called ESG Controversy category score which is computed based on 23 controversy topics such as tax fraud, consumer and environmental controversies or employee safety controversies. These negative events are collected when reported in the media. The list of controversy measures can be found in Table B.2 in Appendix B. The Controversy category score can be interpreted in the following way: the lower the value, the more controversies the company has undergone in the particular year. Lastly, the overlay of the ESG scores data and the Controversy category score is available in the data set. This score is denoted as ESGC score and its primary goal is to reflect the controversies (negative media coverage) in the ESG performance. The ESGC therefore incorporates the information on both ESG and controversies of a particular company.

The calculation of the scores is based on a percentile procedure (unlike the standardized z-score in the ASSET4 database). This approach reflects how many companies have a worse value than the current one, how many have the same value and how many have a value at all. This database is also solving the issue that some industries might have higher average ESG scores than others (then a so called best-in-class screening usually needs to be done to compare how company is performing relative to its industry peers). Here industry benchmarks are used to account for the differences. The Thomson Reuters Business Classification (TRBC) Industry Groups are considered for the calculation of the Environmental and Social scores. For Governance score, country benchmarks are used as the governance principles are usually relatively consistent within countries. Finally, to compute the overall score, weights are used for categories according to the number of measures constituting them thus being a proxy for their importance. Such method challenges the equal weighting scheme in ASSET4 where the Environmental, Social and Governance pillars all have the same importance.

In summary, according to Thomson Reuters the ESG Scores from EIKON are an enhancement and replacement to the existing equally weighted ASSET4 ratings. Most important are the following improvements: (1) the overlay of ESG score with the company controversies, (2) weights depending on the extent of each category, (3) scores adjusted for industry and country benchmarks and (4) the method of percentile rank scoring.

For the reasons stated above, Thomson Reuters ESG Scores from EIKON database will be used. The database was provided by Thomson Reuters solely for the purposes of this thesis in June 2017. Importantly, I use the whole database, including both active and delisted companies and therefore overcoming any survivorship bias that would arise otherwise. Such raw sample contains ESG information on $6,219$ public companies from which $2,114$ are U.S. companies. The companies are identified by a company name and an ISIN code of its equity.

The ESG dataset that I obtained from Thomson Reuters contains the ESG, ESG Controversy score and ESGC overlay score together with the 10 categories (3 environmental, 4 social and 3 governance). These categories weights for the total ESG depend on number of indicators in the particular category (Thomson Reuters, 2017). I verify that the total score is the weighted average of the 10 categories. Then I construct the score for the environmental (ENV), social (SOC) and governance (GOV) pillar with weights depending on number of indicators for the categories in the pillar. The weights for the total ESG score as well as for the pillar scores can be seen in Table 3.1. For example, the 'Emissions' category comprises of 22 indicators and has the weight $w_{ESG} = 22/178 = 12.4\%$ in the total ESG score and $w_{ENV} = 22/(20 + 22 + 19) = 36.1\%$ in the ENV score.

Table 3.1: Weights of categories in the ESG and pillar scores

Pillar	Category	Indicators	Total ESG weights	Pillar weights
ENV	Resource Use	20	11.2%	32.8%
	Emissions	22	12.4%	36.1%
	Innovation	19	10.7%	31.1%
SOC	Workforce	29	16.3%	46.0%
	Human Rights	8	4.5%	12.7%
	Community	14	7.9%	22.2%
	Product Responsibility	12	6.7%	19.0%
GOV	Management	34	19.1%	63.0%
	Shareholders	12	6.7%	22.2%
	CSR strategy	8	4.5%	14.8%
Total		178	100.0%	

Lastly, I also verify that the ESGC score is the average of the ESG and ESG Controversy score. For example, Deutsche Bank's total ESG score in 2009 was 0.86. However, the ESG Controversy score is just 0.01. Then the total ESGC score is 0.44, which lowers the ESG score significantly. After this I end up with five annual panel data of scores (ESG, ESGC, ENV, SOC, GOV) which will then be used as a trading signal for the following year's buying/selling decision.

Table 3.2 shows the descriptive statistics of the Thomson Reuters ESG scores. The overall ESG mean score is slightly lower for the U.S. panel than the global database. The ESGC mean is 0.0324 lower for the U.S., which indicates more controversies among U.S. firms and therefore lower level of the ESGC score. Notably, the environmental score ENV is 0.0414 lower in the U.S. as compared to the global database. This suggests that the U.S. companies are inferior to the international sample when considering environmental indicators. The U.S. sample performs relatively the same on social characteristics and a bit higher in governance than the global universe. As expected, the U.S. sample has lower variation in the cross-section than the global database and about the same in the time-series (apart from GOV which has larger spread in the U.S.). Considering the size of the database, the U.S. sample makes 30.34% of the global ESG scores.

Even though the governance scores of the Thomson Reuters ESG database are supposed to be adjusted for country differences, it has been found that the level of CSR differs depending on the country's legal origin even after the corporate governance issues have been accounted for Liang and Renneboog (2017). This implies that the overall score is still correlated with the type of law in the country and therefore an analysis on a global sample would be biased. From now on the further analysis will be focused on the U.S. sample.

Table 3.2: Descriptive statistics: Thomson Reuters ESG database

This table presents the descriptive statistics for the U.S. sample and the global (raw) database. The means, standard deviations, minima and maxima are reported for the ESG, ESG Controversy overlay, environmental, social and governance scores panel data. *Overall* variables are connected with the full dataset, *between* with the cross-section and *within* with the time-series dimension. Minima and maxima are computed based on variable $x_{i,t}$ for the overall, $\bar{x}_i$ for between and $x_{i,t} - \bar{x}_i + \bar{x}$ for within scale. The panel comprises of the total of 45,891 (13,923) observations for 6,219 (2114) companies in the global (United States) sample over an average time-series of $\bar{T}$ years.

		US sample					Global sample				
Variable		Mean	SD	Min	Max	Observations	Mean	SD	Min	Max	Observations
ESG	overall	0.4860	0.1684	0.0782	0.9808	$N = 13,923$	0.4989	0.1737	0.0599	0.9808	$N = 45,891$
	between		0.1383	0.0998	0.8795	$n = 2,114$		0.1520	0.0970	0.9140	$n = 6,219$
	within		0.0844	-0.1748	0.8633	$\bar{T} = 6.59$		0.0816	-0.1619	0.8947	$\bar{T} = 7.38$
ESGC	overall	0.4139	0.1395	0.0782	0.9567	$N = 13,923$	0.4463	0.1564	0.0599	0.9567	$N = 45,891$
	between		0.1046	0.0998	0.7847	$n = 2,114$		0.1270	0.0970	0.8757	$n = 6,219$
	within		0.0961	-0.0489	0.8687	$\bar{T} = 6.59$		0.0969	-0.0165	0.9144	$\bar{T} = 7.38$
ENV	overall	0.4569	0.2141	0.0269	0.9905	$N = 13,923$	0.4983	0.2228	0.0249	0.9947	$N = 45,891$
	between		0.1733	0.0485	0.9556	$n = 2,114$		0.1939	0.0485	0.9556	$n = 6,219$
	within		0.1123	-0.2085	0.9071	$\bar{T} = 6.59$		0.1077	-0.1671	0.9586	$\bar{T} = 7.38$
SOC	overall	0.4973	0.1933	0.0408	0.9895	$N = 13,923$	0.4995	0.2119	0.0314	0.9940	$N = 45,891$
	between		0.1576	0.0612	0.9502	$n = 2,114$		0.1828	0.0488	0.9502	$n = 6,219$
	within		0.1053	-0.1752	0.9508	$\bar{T} = 6.59$		0.1051	-0.1730	0.9805	$\bar{T} = 7.38$
GOV	overall	0.5036	0.2180	0.0263	0.9934	$N = 13,923$	0.4988	0.2150	0.0091	0.9934	$N = 45,891$
	between		0.1853	0.0535	0.9248	$n = 2,114$		0.1825	0.0168	0.9715	$n = 6,219$
	within		0.1272	-0.1889	0.9961	$\bar{T} = 6.59$		0.1277	-0.1936	1.0628	$\bar{T} = 7.38$

Next, I download industry information for the stocks to continue with the negative screening later in the analysis. Thomson Reuters Business Classification (TRBC) is obtained from Datastream for the stocks in the sample. This is a classification framework that gives information on 5 levels: 10 economic sectors, 28 business sectors, 54 industry groups, 136 industries, and 837 activities. The information on activities and their description will be used. For this study I classify a stock as 'sin' in case it is associated with one of the following activities: alcohol, tobacco, gambling, oil, weapons, nuclear, or biotechnology. Table 3.3 shows the number of sin stocks in the sample according to the sin activities. For the non sin companies (also called the 'no harm' sample), the economic sectors are reported.

Substantial amount of sin in the sample comes from oil-related activities. Sin activities such as alcohol or tobacco have a relatively low representation in this sample with only 6 and 7 companies, respectively. No nuclear or weapon related companies were found in the sample. Rest of the companies in the sample come mostly from Financials.

Table 3.3: Industry breakdown of the sample

sin		no harm			
Oil	124	Technology	291	Consumer Non-Cyclicals	111
Biotech	71	Basic Materials	121	Consumer Cyclicals	292
Gaming	15	Industrials	292	Financials	451
Tobacco	7	Utilities	84	Energy	11
Alcohol	6	Healthcare	151	Telecommunication Services	25

The natural level of the ESG scores across industries differs, however, the cross industry analysis here does not represent a problem because the Thomson Reuters ESG scores are adjusted for industry differences.

Table 3.4 shows the descriptive statistics of the U.S. ESG panel separately for the sin and no harm sample. The mean ESG score for the sin sample is 0.0041 lower than the no-harm sample. However, when accounting for controversies (ESGC score) the mean score is approximately the same for both samples. Therefore, it cannot be claimed that on average the sin companies are more controversial than the non-sin companies with respect to the score level. The mean ENV score is even slightly higher for the sin sample, however the mean SOC score is lower meaning that on average, sin companies are inferior to the no-harm companies considering the social indicators. The mean GOV score is comparable for both samples.

Table 3.4: Descriptive statistics: Scores in sin and no harm sample

This table presents the descriptive statistics for the U.S. sample of sin and no harm companies. The means, standard deviations, minima and maxima are reported for the ESG, ESG Controversy overlay, environmental, social and governance scores panel data. *Overall* variables are connected with the full dataset, *between* with the cross-section and *within* with the time-series dimension. Minima and maxima are computed based on variable $x_{i,t}$ for the overall, $\bar{x}_i$ for between and $x_{i,t} - \bar{x}_i + \bar{x}$ for within scale. The panel comprises of the total of 12,463 (1,460) observations for 1,891 (223) companies in the no harm (sin) sample over an average time-series of 6.59 years.

		US sin sample					US no harm sample				
Variable		Mean	SD	Min	Max	Observations	Mean	SD	Min	Max	Observations
ESG	overall	0.4823	0.1625	0.1452	0.9225	N = 1,460	0.4864	0.1691	0.0782	0.9808	N = 12,463
	between		0.1272	0.1769	0.8219	n = 223		0.1396	0.0998	0.8795	n = 1,891
	within		0.0830	0.1640	0.8250	$\bar{T}$ = 6.59		0.0846	-0.1743	0.8637	$\bar{T}$ = 6.59
ESGC	overall	0.4132	0.1323	0.1351	0.8824	N = 1,460	0.4139	0.1403	0.0782	0.9567	N = 12,463
	between		0.0936	0.1654	0.6913	n = 223		0.1058	0.0998	0.7847	n = 1,891
	within		0.0958	0.0969	0.8400	$\bar{T}$ = 6.59		0.0962	-0.0489	0.8687	$\bar{T}$ = 6.59
ENV	overall	0.4630	0.1968	0.0861	0.9905	N = 1,460	0.4562	0.2161	0.0269	0.9873	N = 12,463
	between		0.1515	0.0945	0.9170	n = 223		0.1756	0.0485	0.9556	n = 1,891
	within		0.1073	0.0310	0.8146	$\bar{T}$ = 6.59		0.1129	-0.2092	0.9064	$\bar{T}$ = 6.59
SOC	overall	0.4835	0.1881	0.0726	0.9727	N = 1,460	0.4989	0.1938	0.0408	0.9895	N = 12,463
	between		0.1470	0.1162	0.8496	n = 223		0.1588	0.0612	0.9502	n = 1,891
	within		0.1066	-0.0399	0.8317	$\bar{T}$ = 6.59		0.1052	-0.1736	0.9524	$\bar{T}$ = 6.59
GOV	overall	0.5015	0.2171	0.0411	0.9693	N = 1,460	0.5038	0.2181	0.0263	0.9934	N = 12,463
	between		0.1836	0.0797	0.8812	n = 223		0.1855	0.0535	0.9248	n = 1,891
	within		0.1211	0.0090	0.8423	$\bar{T}$ = 6.59		0.1279	-0.1886	0.9963	$\bar{T}$ = 6.59

Company ISIN codes obtained together with the ESG database are then used to download static and time series company information from Thomson Reuters Datastream. The time series data is retrieved on a monthly basis for the period from December 2002 to May 2017. As is commonly known, the Datastream data exhibit certain inconveniences that need to be dealt with. Here I follow Ince and Porter (2006) who describe these issues and suggests steps to be taken. The data is therefore screened before it is used for further computations. These screens include static as well as dynamic (time series) screens. Next, I will describe the corresponding data filters that ensure the reliability of the data used for subsequent analysis.

Firstly, I check the set of the ISIN codes by downloading a time series of stock prices. Here the attention is paid to companies where ISIN code was not matched and no price data is available for the security. This concerns 48 companies which are examined separately[1]. One of the reasons for the missing price was found to be a change of a name of a company. In those cases, the correct ISIN was matched. In the rest of the cases, a wrong type of security was matched by the ISIN. I check these securities and find that the ISIN has matched an OTC equity rather than an exchange-traded stock. This concerns 40 companies and I assign the appropriate ISIN. After this, all ISIN codes are checked again by downloading the Stock type to make sure that the instruments are now only of the equity type. Because the ISIN code is connected with a security and all shares listed on different exchanges have the same ISIN, it needs to be checked whether the equity is the primary listing. The Datastream variable Quote indicator is downloaded for this purpose. All stocks are of a primary listing. Other static variables obtained for the stocks in the sample are Geographical classification of company (listing country of a security), Equities status - active, dead, or suspended, and Inactive date[2].

Secondly, the time series screening is conducted. To download the time series information, the problem of delisted companies needs to be solved. My data set includes also companies which are not traded anymore to avoid survivorship bias in the analysis. From each company I use the indicator on whether it is still actively traded or not, denoted by ACT. and DEAD, respectively. The full sample contains 1745 active and 369 dead U.S. companies. The dead companies are not active anymore due to a delisting, a merger, or an acquisition.

For the period after a company is delisted, the Datastream returns the last price available all the way until the end date of the sample (May 2017). If not accounted for, the issue would lead to zero returns even though the company is not listed and cannot be traded on an exchange any more. One possibility is to delete the repeating price at the

[1] The Datastream Navigator is used to search for company information such as identifiers, instrument types, stock price availability, status, or company description

[2] Datastream codes for the static variables are: TYPE (Stock type), GEOGN (Geografical classification), ESTAT (Equities status), WC07051 (Inactive date), ISINID (Quote indicator).

end. Ince and Porter (2006) use this method but note that this might remove a number of valid constant price data points just before the company's delisting date and thus some observations of valid zero returns may be excluded. Therefore, I use the download settings in Datastream to request that the price observations are not taken by the last price available. In this way, the data will be missing for this time periods and the dead company will not be included in the portfolio at a later stage of the analysis.

For all the equities I extract unadjusted prices (UP mnemonic) as well as total return indexes (RI mnemonic). In the Datastream, the data type mnemonic X(UP#T)˜U$ and (X(RI)˜U$)*(X(P#T)/X(P#T)) is used to suppress padding of the price and total return index since the time the stock has been inactive. The total return index shows the growth in value of a share assuming the re-investment of dividends. The returns are then computed as the change of the index value. Unlike the price returns, which measure only the capital appreciation (price change) of the stock, the total returns measure returns earned including cash dividends and their reinvestment. In general, historical performance of stocks gives a different picture when considering price return and total return. The total returns more precisely represent the returns reaped by the market participant. To examine the equity investments the returns generated from both sources (capital appreciation and cash dividends) will be considered and therefore the total return is used due to the power of compounding of dividends reinvestment.

Apart from padding the value of a dead company, other time series issues require corrections. In the academic literature it is common to remove small stocks (sometimes also called "penny stocks"). This is because the Datastream rounds the prices to the nearest penny and for small prices the calculated returns might be considerably far from the corresponding returns where prices are not rounded. Therefore, I follow the procedure of Ince and Porter (2006) and remove the return observations for which the end-of-previous-month price is less than 1 USD. The next issue in the data that needs to be dealt with is a data error that causes a large return, which is nevertheless reversed within one month. Ince and Porter (2006) set a threshold 300% which they admit to be rather arbitrary but report that it performs well and so it will be taken here as the appropriate level for reversal. Following Ince and Porter (2006), if R_t or R_{t-1} is greater than 300% and $(1 + R_t)(1 + R_{t-1}) - 1$ is less than 50%, I set R_t and R_{t-1} to missing. This affects 30 observations in the sample (15 such reversals have been found). In the following step, all returns above 990% are removed for probable data errors. Lastly, market capitalization (MV mnemonic) is obtained for all the equities and the return data are aligned with the market cap information, i.e. observations for which the previous end of month market value is not available are set to missing.

By this procedure, the data was screened for non-common equity and various data errors. I end up with a sample of monthly U.S. stock returns time series which spans from December 2002 to May 2017, with the returns reflecting a stock's trading life and thus not available for the stock's inactive life. The screens performed on the raw sample are believed to be adequate and the transformed data obtained by this procedure is considered appropriate for the following analysis.

In the following step, I align the ESG and stock data. The idea is to use the ESG information for stock i from year t to trade stock i in year $t+1$. However, this is only possible if the return for stock i is available in year $t+1$. Therefore I check the time coherence of the ESG scores and stock return data. For the stock-month observations where the ESG information is available but the stock return is not, I delete the ESG data point because this information cannot be utilized. As a result, $4,077$ stock-month ESG observations cannot be used as a source of information for trading, which is 2.66% of the ESG stock-month sample. Once the ESG and stock data are aligned, the sample contains $153,069$ stock-month observations from the total panel of 2114 U.S. companies and 173 months.

Figure 3.1: Number of stocks in the sample

This figure shows the number of stocks in the sample over the period January 2003 to May 2017. The solid line represents the full sample and the dashed line represents the sample obtained after removing companies classified as 'sin stocks'.

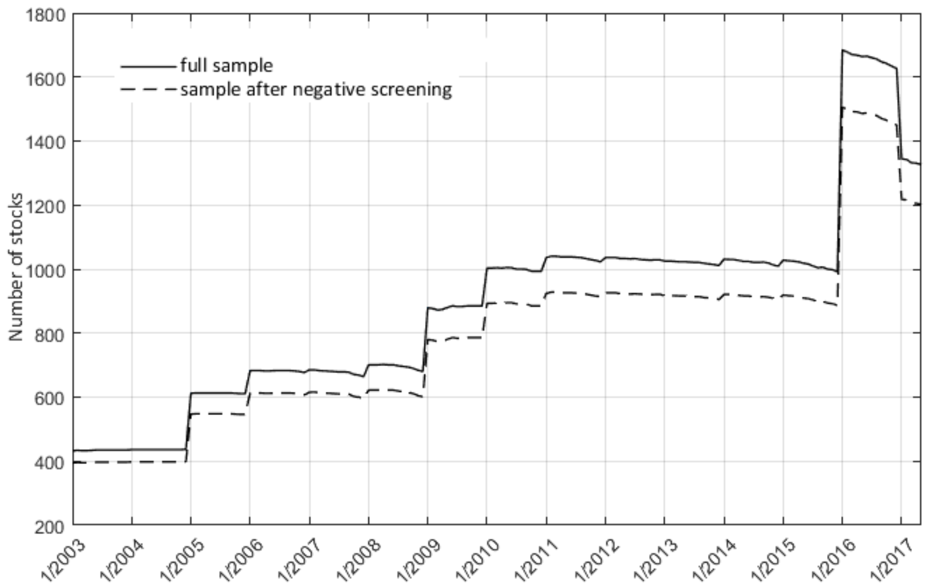


Figure 3.1 shows the number of stocks in the sample during the sample period. The shape of the line reflects the inclusion of further indexes into the ESG database as described in the data section. The most considerable increase is visible at the beginning

of 2016 and is connected with including companies from the Russel 2000 index. The dashed line shows the sample obtained after negative screening, i.e. after 223 companies from 'sin' industries were removed.

For further analysis I also download the Fama French risk factors and momentum, which are available at the Kenneth French's online data library[3]. From this website the risk-free rate, market, size, value and momentum factors are obtained for the sample period. The market return is the return of the CRSP value-weighted index less the risk free rate. These are in form of time series returns in percent format and for further analysis will be divided by 100 to match the format of the stock returns. I download the factor data at time when factors are computed using the CRSP database from August 2017. For the comparison of the stocks in the sample with a broader U.S. universe I download the U.S. Research Breakpoints from the Kenneth French's data library, specifically ME breakpoints for market value and BE/ME breakpoints for book-to-market ratio. These come in form of a percentile computed based on NYSE stocks and list every percentile from the 5% until 100%. To make such comparison possible, I need market value and book-to-market variables for my sample which would match the French's ME and BE/ME by definition.

French defines market value as $ME = p{*}shares\ outstanding$. I cannot use the MV variable from Datastream because it is defined as $MV = p{*}shares\ in\ issue$ and a problem of mismatch would be faced as $shares\ issued = shares\ outstanding{+}treasury\ shares$. Therefore I construct ME for the companies in my sample by multiplying the unadjusted price by number of shares outstanding at month end[4]. French's ME/BE variable is of a yearly frequency and is computed for the end of each June. BE/ME in June of year t is computed as book equity BE from last fiscal year end in $t-1$ divided by market value ME, which is the price times shares outstanding at the end of December of $t-1$. I download the BE for the previous year[5] and divide it by ME on a yearly basis. To reduce the effect of outliers I winsorize both ME and BE/ME at 1% and 99% level. This means that values above the $99th$ percentile and below the $1st$ percentile are replaced by the value of $99th$ and $1st$ percentile, respectively.

The Figure 3.2 depicts the comparison of the sample market value to the French's breakpoints. The sample median and average are shown together with the percentiles of the U.S. market. At a first sight it is clear that the average is significantly higher than the average. Winsorizing at lower level would push the average down, however, the median

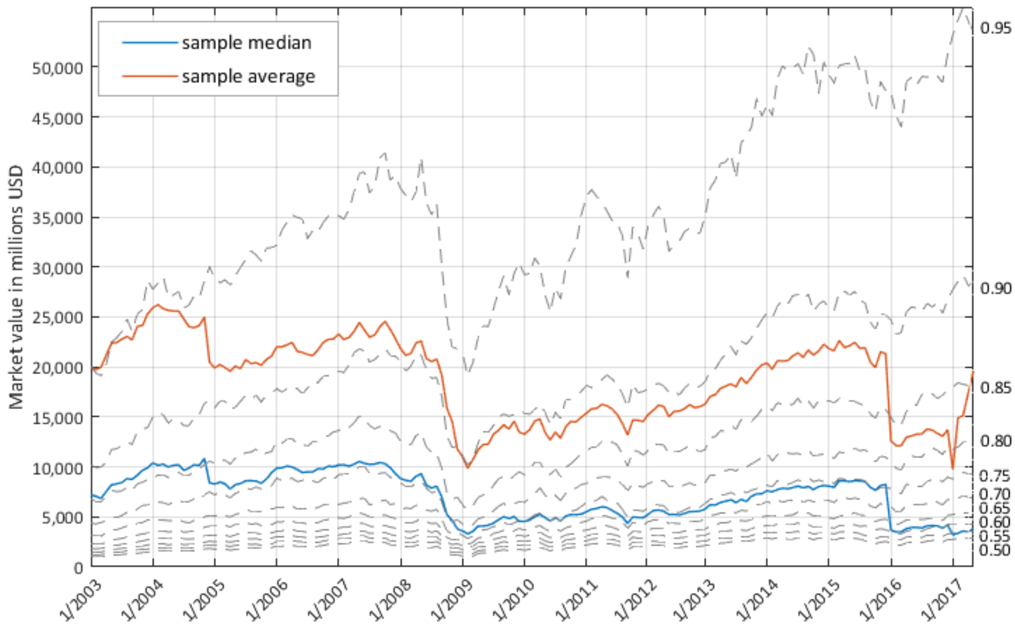
[3]http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[4]This variable is downloaded from Datastream by the code WC05301, it is in thousands and I further divide it by $1,000$ to get a variable in millions, which matches the French definition.

[5]In Datastream it is variable WC05491. It is divided by number of shares outstanding so I multiply it again to get BE defined same way as by French.

Figure 3.2: Market value comparison to the U.S. market

This figure shows the sample market value and the market value percentiles of the broader U.S. market from January 2003 to May 2017. The red (upper) solid line denotes the sample average and the blue (lower) solid line denotes the sample median. From top to bottom, the dashed lines represent the 95*th*, 90*th*, 85*th*, 80*th*, 75*th*, 70*th*, 65*th*, 60*th*, 55*th*, and 50*th* percentile of the U.S. market as provided by the Kenneth French's data library in the U.S. Breakpoints file.



stays the same and therefore I report both at only 1 and 99% winsorization to present the data with less intervention. The sample median starts at 7.2 billion USD, which is just above the U.S. 85*th* percentile. In the period of 2005 to mid 2008 it continues slightly above the 80*th* percentile and then falls sharply in value with the financial crisis. Until the end of 2015 it fluctuates in between the 70*th* and 75*th* percentile and the median increases gradually to 8.2 billion USD at the end of 2015. At the beginning of 2016 the value falls abruptly to the 55*th* percentile, which indicates a sudden change in the market cap characteristics of the sample.

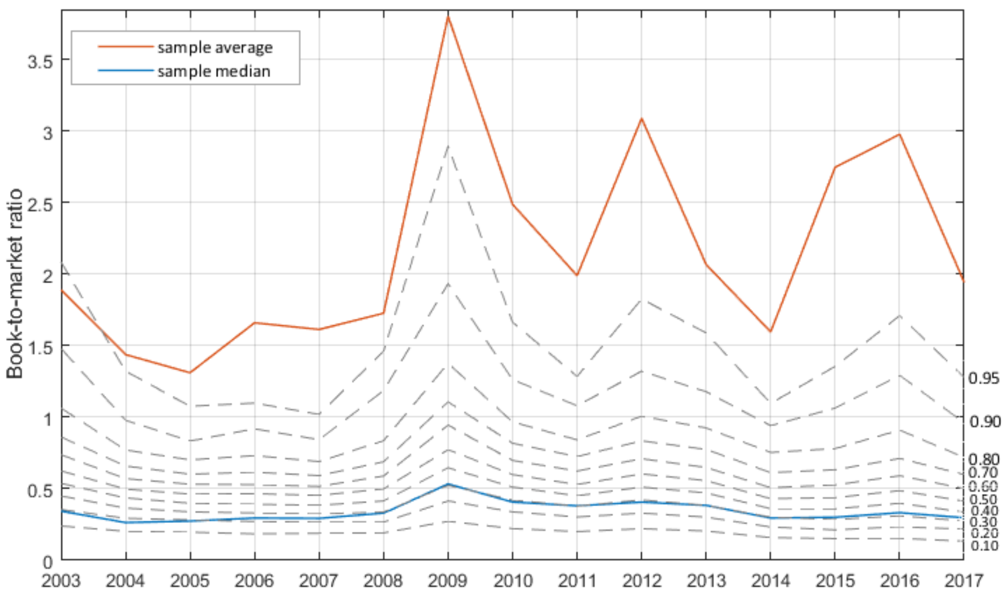
The sample average is considerably higher than the median and suggests that even after the data is winsorized, many extreme values remain in the sample and affect the average. This all indicates that the sample comprises of relatively large companies as compared to the broader U.S. universe. Until 2008, approximately half of the sample were large cap companies (above 10 billion USD). After 2008, the sample follows the market crash with a rapid fall in value. Compared to the U.S. market, the sample value has fallen approximately the same as it copies the 90*th* percentile during the crisis. After the crisis, however, it picks up more slowly than the U.S. market and stays between the

85th and 90th percentile.

The Figure 3.3 depicts the comparison of the sample book-to-market ratio to the French's breakpoints. Again we can see the sample median and average as well as the percentiles of the U.S. market. At the beginning the median follows the 20th percentile. In 2008 it increases to the level of the 30th percentile and copies it until the end of the time period. The average book-to-market is substantially higher. In fact, it rises above the 95th percentile already at the beginning of the period and stays above until the end of the sample period. This indicates that the distribution of B/M ratio of the sample is highly positively skewed. In general, the B/M ratio of the sample follows the U.S. market, including the crash of 2008. The median suggests that more than half of stocks in the sample are overvalued (B/M bellow 1). These growth stocks keep the median B/M low, however as the average suggests, there is a considerable amount of potentially undervalued stocks that keep the average high.

Figure 3.3: Book-to-market ratio comparison to the U.S. market

This figure shows the sample book-to-market ratio (B/M) and the B/M percentiles of the broader U.S. market on a yearly basis from June 2003 to June 2017. The red (upper) solid line denotes the sample average and the blue (lower) solid line denotes the sample median. From top to bottom, the dashed lines represent the 95th, 90th, 80th, 70th, 60th, 50th, 40th, 30th, 20th, and 10th percentile of the U.S. market as provided by the Kenneth French's data library in the U.S. Breakpoints file.



As the last step of acquiring the data for the empirical analysis, the sentiment index is needed. Following Stambaugh et al. (2012) I use the sentiment index created by Baker and Wurgler (2006) (abbreviated as BW index). Another alternative would be to consider the commonly used sentiment index constructed by the University of Michigan Surveys of

Consumers. However, this index reflects opinions of 500 random households on economy and is probably less connected with the sentiment of stock market participants than the BW index (Stambaugh et al., 2012).

The BW index is constructed by combining six sentiment proxies: the closed-end fund discount, NYSE share turnover, the number of average first-day returns on IPOs, the equity share in new issues, and the dividend premium. Each variable reflects a sentiment part as well as idiosyncratic part not related to sentiment (Baker and Wurgler, 2006). The authors use the principal components analysis to isolate the common component. However, it could be argued that apart from sentiment, this index could reflect macroeconomic conditions. Baker and Wurgler (2006) therefore construct an adjusted index that removes the business cycle variation of the sentiment proxies before they enter the principal components analysis. This is done by regressing the sentiment proxies on a set of macro variables such as growth of production, growth of consumption, growth of employment or a dummy for recessions. By such modification, the pure effect of sentiment should be captured.

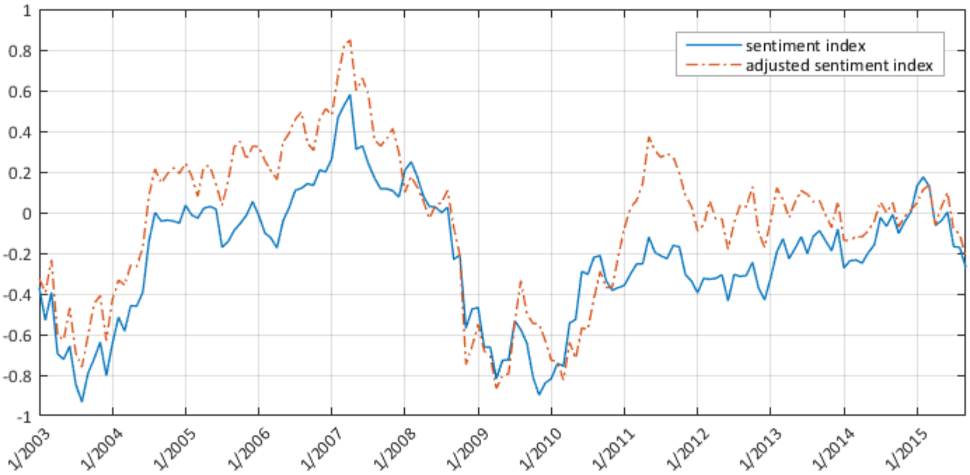
I download the BW index from Jeffrey Wurgler's website[6]. This index is available since July 1965 until September 2015. Figure 3.4 plots the sentiment index since the beginning of the sample period. The figure shows the original index as well as index that is adjusted for a set of macroeconomic variables. The correlation of the two indices over the sample period is 0.90 while the correlation of the two indices over the period of the index data availability is found to be 0.97. This suggests that in the last 13 years of the index data period, the business cycles effects were more pronounced than in the earlier period.

At the beginning of the sample period we can see a drop in the index value, which could still be attributed to the long decline after the Internet bubble peak. After this initial drop, there is a gradual increase in the level of sentiment from mid 2003 to the beginning of 2007. Then the sentiment keeps falling until early 2009 when the level is low as a consequence of the global financial crisis. During 2009 it jumps shortly up but reverses back and then starts rising in 2010. Since then the index level is up but fluctuating during the following 5 years. In 2015 the value decreased again copying the stock market decline. In the periods of 2005-2008 and 2011 - 2014 we can see that the value of the adjusted index is slightly higher than the unadjusted index. This could be explained by the fact that even though the effect of the macroeconomic variables is removed, the sentiment is fueled by beliefs about economic fundamentals and it even overshoots. Then, after the adjustment, we can see higher sentiment level in periods of economic expansion.

[6] http://people.stern.nyu.edu/jwurgler/

Figure 3.4: Sentiment index

This figure shows the sentiment index created by Baker and Wurgler (2006) plotted over the period January 2003 to September 2015. The blue (solid) line represents the original index and the red (dash-dotted) line represents the index adjusted to remove the effect of business cycles.



3.2 Methodology

In this section the research methodology will be presented. First, the formation of ESG portfolios will be explained. Then, the factor model will be described and the methodology to test the sentiment hypothesis will be given. Finally, additional robustness tests will be outlined.

3.2.1 Portfolio construction

From the empirical literature it can be seen that forming portfolios based on the ESG scores is a common standard for examining the link between CSR and financial performance. This method utilizes information from a large panel data set and provides us with a time-series of returns to be evaluated. The portfolio construction based on the ESG scores represents a straightforward trading strategy that will be used in this empirical analysis.

For each month t the stocks are sorted in descending order based on the score available in month $t-1$. Firstly, I construct 10 portfolios that consist of buying stocks in the 10 different deciles of the score-ranked stocks. This should give an indication of difference in mean returns of strategies based on different levels of ESG scores. This is done based on ESG, ESGC, ENV, SOC and GOV scores.

Then, I form portfolios based on a subset of the sample for which the score is above or bellow a certain threshold. This threshold is computed separately in each month and is based on percentile of the score distribution in that month. There is no consensus in the literature on the most appropriate cut-off level. Empirical studies usually perform the analysis using more cut-off levels. For example Kempf and Osthoff (2007) use 10% as the base strategy and then perform additional checks with larger subsets of the sample. More recent studies use stricter threshold, such as Auer and Schuhmacher (2016) who start with 5% or (Halbritter and Dorfleitner, 2015) who even use 1% and then shift the threshold to create larger portfolios. I conduct this study using four cut-off levels: 5%, 10%, 20%, and 30%. The strategy using the 5% threshold is not that reliable because it is based on a small sample (especially at the beginning of the period), yet it will be used to provide us with an insight into the sample with the most extreme ESG ranking. The 10% cut-off strategy will then be considered as the main strategy.

First approach is buying the highest ranked stocks (here denoted as the top portfolio). Second approach is buying the stocks that perform the worst in the ranking (denoted as the bottom portfolio). The last portfolio is the difference of the top and bottom portfolios and simulates holding long position on the best-performers and short position on the worst-performers (denoted as L/S or long-short strategy). Here I suppose that shorting has no transaction costs and no collateral has to be deposited on the margin account. The long-short strategies are denoted by the cut-off that applies to the top and bottom part of the sample, therefore the long-short strategy associated with the 5% cut-off is holding a certain position in 10% of the stocks in the sample.

Again, these portfolios are constructed using the information from ESG, ESGC, ENV, SOC and GOV score. According to which percentile is traded, I will denote the top portfolio strategy by ESG_{t5}, ESG_{t10}, ESG_{t20} and ESG_{t30}; for the bottom portfolio strategy ESG_{b5}, for the long-short strategy ESG_{ls5} and similarly for the other percentiles and score types. Decile and percentile portfolios are constructed using both equal weights and weights determined by market capitalization from the previous month. I repeat the whole procedure on the sample which is screened for harmful industries, i.e. 'sin stocks' are removed and then the portfolio is constructed based on the scores as described above. Further on I will denote this approach as the *strict* ESG, *strict* ESGC, *strict* ENV, *strict* SOC or *strict* GOV strategy. Due to the low number of the sin companies in the sample (especially at the beginning of the sample period), it will not be possible to examine strategies that trade sin stocks based on their ESG ranking. The monthly returns from all strategies will be expressed as returns above the risk-free rate. This will be referred to as excess returns of the particular strategy[7].

3.2.2 Performance evaluation

In the subsequent part, the steps to evaluate the performance of the trading strategies are presented. For this thesis the performance evaluation is set up with three criteria in mind: (i) strategy performance has to be compared to alternatives, (ii) performance needs to be evaluated by suitable measures and (iii) the ability to select stocks needs to be distinguished from random luck.

At the beginning of this section, return of one exemplary strategy will be taken to examine the properties of the return series. The time series process will be analysed through a simple plot over the sample period as well as the sample autocorrelation function. The basic properties of the time series process such as stationarity and autocorrelation will be commented on.

To start with basic evaluation, annualized mean excess returns will be computed as $r_{annualized} = (1+r)^{12} - 1$, where r is the mean monthly excess return for the particular strategy. The next performance indicator is the Sharpe ratio, which measures the trade-off between the excess return and total volatility. It accounts for the total risk, i.e. both the systematic and unsystematic component. It is constructed in a way that it assumes that the portfolio represents the investor's whole investment fund. The definition by Jobson and Korkie (1981) will be used, which defines the Sharpe ratio as $\hat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}$, where $\hat{\mu}$ is the sample mean of the returns above the risk-free rate $\hat{\mu} = \frac{\sum_{t=1}^{T} R_t - R_{f,t}}{T}$; and $\hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{T}(R_t - R_{ft} - \hat{\mu})^2}{T-1}}$ is the sample standard deviation of excess returns. Then the Sharpe ratio is annualized to be reported next to the annualized mean returns: $SR_{annulaized} = SR * \sqrt{12}$. A higher Sharpe ratio is associated with a higher t-statistic and therefore a higher significance level for the strategy (Harvey and Liu, 2015). The Sharpe ratio is a commonly used evaluation measure in the ESG-related literature, however, it serves merely as a first step in the performance evaluation procedure.

There is a highly substantiated critique to making inference based only on comparing the Sharpe ratios of different strategies. Such Sharpe ratio is just an estimate and needs to be tested. Even though there are methods to test the Sharpe ratio such as the test developed by Jobson and Korkie (1981), these are seldom used in the ESG-related literature. This test is in fact not valid when performed on non-normally distributed

[7]Throughout this paper, the term 'excess returns' is associated with the returns above the risk-free rate while the term 'abnormal returns' is connected with the returns left after adjusting for (risk) factors, also denoted as Alpha in the factor models.

returns (e.g. in presence of serial correlation). Moreover, Lo (2002) describes that the time aggregation for q-period Sharpe ratio $SR(q) = \sqrt{q}SR$ can be computed also only for independent identically distributed returns. The author states that the Sharpe ratio should not be used as the only evaluation criterion. Harvey and Liu (2014) confirm that the Sharpe ratio may be misleading despite its widespread use. They also claim that the Sharpe ratio is likely to be inflated.

Even if we disregard the disadvantages of the Sharpe ratios mentioned above, the reason to use a better evaluation measure is that the Sharpe ratio reflects both the systematic and non-systematic risk while the aim is to risk-adjust the returns only for the systematic part of risk because it cannot be avoided by diversification. An alternative measure could be the Treynor ratio, which uses only the systematic risk to account for the level of risk. However, the standard in the academic ESG-related literature is to proceed with a factor model to risk-adjust the mean excess return. For the reasons stated above, I will report the annualized Sharpe ratio as a performance indication of the strategies, however, I will consider the factor model as a more appropriate evaluation tool.

To choose the performance benchmarks I follow the standard in the ESG literature. Firstly, the capital asset pricing model (CAPM) will be estimated, where the only factor is the market excess return. The estimated time-series regression is

$$R_t - R_{f,t} = \alpha + \beta(Rm_t - R_{f,t}) + u_t \tag{3.1}$$

where R_t are the portfolio returns, Rm_t is the market return and $R_{f,t}$ is the risk-free rate at month t. The intercept is the Jensen's α.

Secondly, the Carhart (1997) 4-factor model will be employed. It is nowadays a conventional performance benchmark and is used as an active management evaluation model (Bodie et al., 2014). The model is estimated by the following equation:

$$R_t - R_{f,t} = \alpha + \beta(Rm_t - R_{f,t}) + sSMB_t + hHML_t + wWMLt + u_t \tag{3.2}$$

where the dependent variable is the excess return on portfolio in month t. SMB_t is the difference between returns of small stocks and big stocks, HML_t is the difference between returns of high book-to-market and low book-to-market stocks and WML_t is the difference between stocks with high prior returns and low prior returns. The abnormal return of the portfolio is expressed by the estimate of α. The reason to include the additional factors is to isolate the differences in returns that are attributable to the ESG-characteristic from the differences in the factor portfolios characteristics. The standards errors will be estimated by the Newey and West (1987) method, which is a common approach in the academic literature. It produces standard errors where the error structure is assumed to be heteroskedastic and possibly autocorrelated across observations up to a certain lag. I choose the lag 12 based on the monthly frequency of the data to account for autocorrelation within the time series. The adjusted R^2, which adjusts for the number of the terms in the model, will be reported to assess the explanatory power of the model. This model will be estimated for all the time series of returns of constructed portfolios with the interpretation focused on the long-short investment strategies.

3.2.3 Sentiment hypothesis testing

In this part I am going to present the methodology to test the hypothesis whether the abnormal returns from the constructed investment strategies could be explained by sentiment-driven mispricing.

For further analysis the sample period is cut for two reasons: (A) The Baker and Wurgler sentiment index is available for download only until September 2015. (B) Based on the sample inspection we could see that the sample characteristics change with the beginning of 2016 (due to inclusion of another index, and only partial availability of scores for 2016 and therefore reduced trading sample in 2017). This will also serve as a check for the results obtained by using the full time span. In the following method I follow Stambaugh et al. (2012) to test the sentiment hypothesis for the ESG anomaly the same way the authors did for 11 anomalies of their choice. I will select suitable long-short strategies to test this hypothesis on after the results from the 4-factor model.

I choose to do this analysis using the adjusted Baker and Wurgler index as it should be cleaned from business cycle effects and therefore simply capture the market sentiment. I denote each month as period of high (low) sentiment if the value of the index is above (below) the median of the sample period.

In the first step I will compute the mean excess returns separately for the months following high-sentiment periods and for those following low-sentiment periods. I will test whether the mean is significantly different from zero and report t-statistics based on Newey and West (1987) standard errors. This should give the first indication whether the anomaly is more pronounced following a high sentiment rather than a low sentiment period.

In the second step I will evaluate the risk-adjusted returns. Unlike Stambaugh et al. (2012) who uses the 3 factor model of Fama and French (1993), I keep the consistency with my previous analysis and use the 4-factor model adding momentum. I modify the 4-factor model to look separately at the abnormal return arising in the months after high and low sentiment periods. This will be done by getting the estimates of a_H and a_L in the regression

$$R_t - R_{f,t} = a_H d_{H,t} + a_L d_{L,t} + b(Rm_t - R_{f,t}) + sSMB_t + hHML_t + wWML_t + u_t, \quad (3.3)$$

where R_t is the return of the investment strategy in month t, $d_{H,t}$ and $d_{L,t}$ are dummy variables taking the value of one if the month follows a high and low sentiment period, respectively. The rest of the factors are defined the same as in the previous 4-factor model.

The first and second step indicate whether the sentiment can be related to the ESG anomaly and therefore whether the factor model could be enhanced by the sentiment index. For the further regressions where the level of the sentiment index is used, the index is re-scaled to have zero mean and standard deviation of 1. In the third stage the excess returns are regressed on the level of the lagged sentiment index:

$$R_t - R_{f,t} = a + bBW_{t-1} + u_t, \quad (3.4)$$

where R_t is the strategy return in month t and BW_{t-1} is the value of sentiment index in month $t-1$.

The final part of this hypothesis testing is adding the sentiment index to the factor model

$$R_t - R_{f,t} = a + bBW_{t-1} + c(Rm_t - R_{f,t}) + sSMB_t + hHML_t + wWML_t + u_t, \quad (3.5)$$

where R_t is the portfolio return in month t and BW_{t-1} is the level of the sentiment index in month $t-1$.

3.2.4 Robustness tests

Apart from conducting the analysis with different portfolio cut-offs and weighting schemes as described above, further robustness tests will be done to possibly help explain the results obtained. Firstly, the 4-factor model will be re-estimated for a shorter period excluding months following December 2015. The reason for this is demonstrated mainly in Figure 3.1 and Figure 3.2. There we could see a dramatic change in the characteristics of the sample. This stems from the enlargement of the database as well as the fact that not all 2016 score were available in the database at the time of analysis to give a trading signal for the following months.

Secondly, I will re-estimate the 4-factor model for two subperiods: January 2003 to

June 2009 and July 2009 to December 2015. The results would indicate whether there is a shift in abnormal returns over time.

Lastly, I will employ the five-factor model of Fama and French (2015) to see the exposure to the newly added quality factors: profitability and investment. This model has not been used yet in the academic literature on ESG investing. The inclusion to this study was mainly motivated by the recent results of Blitz and Fabozzi (2017) who claim to resolve the sin stock anomaly when using this model. Profitability and investment factors are to be appended to the 3-factor model:

$$R_t - R_{f,t} = a + b(Rm_t - R_{f,t}) + sSMB_t + hHML_t + rRMW_t + cCMA_t + u_t, \quad (3.6)$$

where RMW_t is the difference between returns of stocks with robust and weak profitability and CMA_t is the difference between returns of stocks with low and high investment companies (here called conservative and aggressive).

Chapter 4

Empirical Results

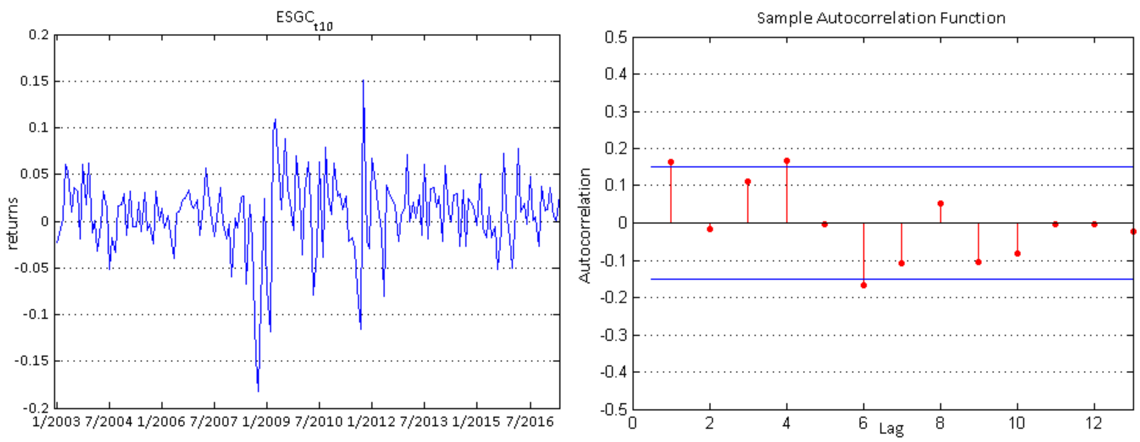
In this chapter the results from the empirical analysis will be presented. First, results from the trading strategies will be described. Then the risk adjusted performance will be shown and the outcome from testing the sentiment hypothesis will be presented. Finally, the results of the robustness tests will be presented.

4.1 Trading strategies

After obtaining the return series from all trading strategies, the strategy $ESGC_{t10}$ is chosen to examine properties of the return series.

Figure 4.1: Return series properties

On the left-hand side the figure shows the monthly excess return series of the strategy buying top 10% of the stocks ranked on ESGC over the time period from January 2003 to May 2017. On the right-hand side, the plot shows the sample autocorrelation function of the same series together with the 95% confidence intervals.

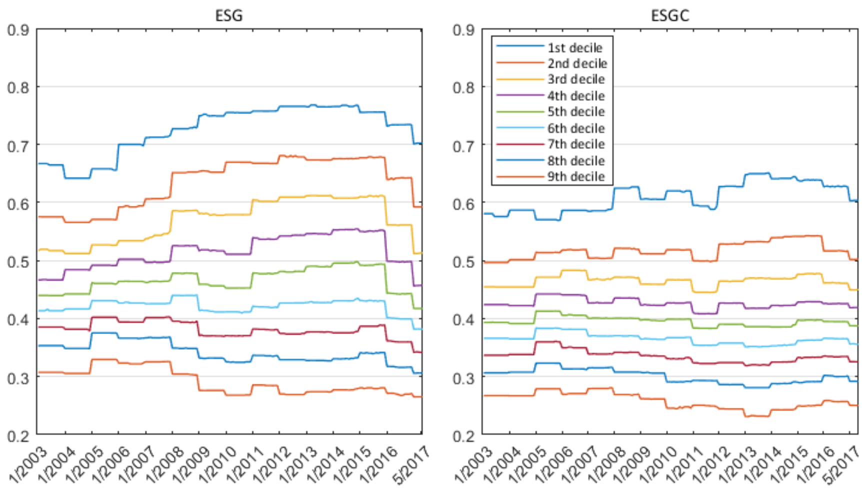


In Figure 4.1 we can see the excess returns plotted over the sample time period.

The mean of the observed time series seems relatively stable over time, however, we might suspect that the volatility changes as the plot seems to have higher variation approximately in the second third of the sample as compared to the first and the last third. The right-hand side of the figure displays the sample autocorrelation function. I test for the autocorrelation in the series by the Ljung-Box test. The null hypothesis is that the first m autocorrelations are jointly zero. The test rejects the null hypothesis $m = 12$ with p-value 0.0283, which suggests that there is a certain degree of linear dependence in the data. The heteroskedasticity of the series cannot be tested by standard tests (e.g. Breusch-Pagan test), because they assume independence of the errors. I test for the unit root of the series by the Augmented Dickey Fuller test, which rejects the null hypothesis of unit root with p-value of 0.001. The series is therefore considered to be stationary. The time series of returns of other strategies exhibit similar properties with slightly different p-values of these tests but the same implications apply. It is therefore deemed appropriate to continue with the analysis of the series while heteroskedasticity- and autocorrelation-consistent standard errors will be used as outlined in the methodology section.

Figure 4.2: The scores of the decile portfolios

This figure shows the ESG and ESGC scores that are the threshold values of the decile portfolios. The topmost line is the lower threshold connected with the top 10% scores and the lowest line is the upper threshold for the bottom 10% scores. The scores are plotted over the sample period January 2003 to May 2017.



Now I will continue by reporting the results of the deciles portfolios. Figure 4.2 depicts the thresholds between particular deciles of the ESG and ESGC scores. For example, the topmost line represents the value above which the stocks fall into the first decile portfolio and the stocks with a value below fall into the second decile portfolio. The fifth decile is the median score of all the stocks in the sample. We can see that in general the ESGC levels are lower than ESG. Notable is also the change over time. In second half of the sample the scores seem to be more spread out. This means that the uppermost decile portfolio is composed of stocks with higher scores and the lowest portfolio includes stocks with worse scores than in the first half of the sample period. This indicates that the U.S. sample ESG characteristics are changing over time, especially that the U.S. sample has a higher variation in the scores later on.

Table 4.1 shows the results from trading strategies of the decile portfolios. The ESG velu-weighted portfolios exhibit annualized mean excess returns from 7.93% to 11.67%, where the best peformance is shown by the third and fourth portfolio. However, when the mean returns are adjusted for risk (expressed by the standard deviation), the Sharpe ratios show the highest value for the lowest decile portfolio. For the other scores, well performing portfolios can be found randomly within the deciles (e.g. for ESGC the 6th decile, for ENV the 4th decile, for SOC the 3rd decile). However, similar pattern appears with the most bottom portfolio outperforming the others based on simple comparison of the Sharpe ratios. These portfolios with higher Sharpe ratio have statistically significant mean returns with p-value bellow 5%. In the equally weighted portfolios, the level of the mean returns is higher, for ESG it is in the range between 10.95% and 15.72%. The Sharpe ratios are also slightly higher. The fact that the equally weighted portfolios exhibit better performance than the value-weighted portfolios indicates that the results of the equal-weighted strategies might be driven by small less liquid stocks. The results of the decile stretegies for portfolios with negative screening (strict strategies) can be found in Table A.1 in Appendix A. The results are in line with the simple strategies with the highest Sharpe ratios and high statistical significance (p-value bellow 1% of the lowest decile portfolio.

Next, I will report the results from the trading strategies investing in a certain percentile of the ESG score distribution. Figure 4.3 shows the size of these portfolios regarding the number of stocks traded. Both the full sample and the strict strategy (negative screening) alternatives are displayed. At the beginning of the sample period, 24 stocks are traded in the upper as well as in the lower 5% of the sample and 22 in the strict strategy, 43 stocks in the 10% subsets of the sample (40 in strict), 86 stocks in the 20% subsets of the sample (79 in strict) and 130 stocks in the 30% subsets of the sample (119 in strict). While these numbers increase slowly for the extreme 5% of the sample, the growth is stronger for the larger subsets. At the beginning of 2011, the 5% sample consists of 52 stocks (46 in strict), the 10% one of 104 stocks (93 in strict) and above 186 and 279 for the strict strategy in the 20% and 30% sample, respectively. In 2016 there is a sharp increase in the sample sizes as more companies were included into the ESG database. The decline towards the end again confirms that not all 2016 scores were

Table 4.1: Decile portfolios

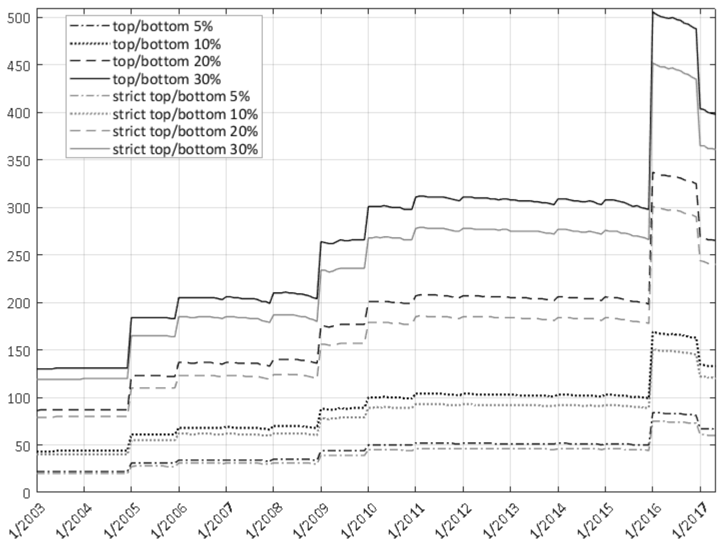
This table reports the annualized mean excess returns and Sharpe ratios (SR) of the decile portfolios formed with equal weights and weights based on market capitalization (value-weighted). The t-statistic comes from testing the excess return series for a null hypothesis of zero mean using the Newey-West standard errors. Decile 1 denotes the top decile (portfolio formed from the stocks with the best scores) and decile 10 denotes the bottom decile (stocks with the lowest scores).

decile		1	2	3	4	5	6	7	8	9	10
						value-weighted					
ESG	mean	0.0793	0.0977	0.1167	0.1104	0.0939	0.0900	0.0953	0.1029	0.0866	0.0929
	t-stat	2.00	2.21	2.45	2.37	1.78	1.77	1.74	2.08	1.59	2.56
	SR	0.62	0.70	0.76	0.73	0.57	0.57	0.59	0.63	0.53	0.86
ESGC	mean	0.0886	0.0997	0.0794	0.1042	0.0607	0.1145	0.1090	0.1068	0.1054	0.0913
	t-stat	1.98	2.18	1.68	2.61	1.26	2.62	2.3	2.08	2.31	2.57
	SR	0.61	0.67	0.56	0.79	0.40	0.81	0.71	0.67	0.68	0.84
ENV	mean	0.0820	0.0909	0.1019	0.1326	0.0933	0.0868	0.0818	0.0946	0.1130	0.0931
	t-stat	1.92	2.13	2.74	2.85	1.94	1.73	1.53	1.76	2.05	2.53
	SR	0.59	0.70	0.77	0.83	0.61	0.56	0.50	0.56	0.74	0.86
SOC	mean	0.0818	0.0843	0.1202	0.0988	0.1005	0.0963	0.1151	0.0984	0.1151	0.0881
	t-stat	1.99	1.83	2.83	2.13	2.01	1.94	2.14	1.75	2.44	2.44
	SR	0.62	0.61	0.83	0.66	0.65	0.63	0.70	0.59	0.76	0.81
GOV	mean	0.0895	0.1001	0.0923	0.0945	0.0951	0.1092	0.1015	0.0740	0.1046	0.0923
	t-stat	2.07	2.4	1.96	2.21	2.04	2.35	1.96	1.52	1.93	2.5
	SR	0.67	0.78	0.62	0.66	0.63	0.69	0.62	0.46	0.63	0.85
						equal-weighted					
ESG	mean	0.1210	0.1419	0.1447	0.1572	0.1171	0.1330	0.1232	0.1367	0.1095	0.1168
	t-stat	2.49	2.71	2.55	2.81	1.92	2.43	2	2.34	1.96	3.37
	SR	0.79	0.87	0.77	0.87	0.61	0.71	0.66	0.72	0.62	1.04
ESGC	mean	0.1383	0.1483	0.1337	0.1379	0.1175	0.1617	0.1214	0.1137	0.1200	0.1167
	t-stat	2.59	2.8	2.36	2.49	2	2.81	2.14	2.01	2.04	3.37
	SR	0.80	0.79	0.78	0.78	0.63	0.90	0.66	0.64	0.65	1.04
ENV	mean	0.1289	0.1357	0.1441	0.1632	0.1265	0.1050	0.1194	0.1285	0.1338	0.1178
	t-stat	2.4	2.75	2.69	2.77	2.29	1.88	2.06	2.11	2.19	3.4
	SR	0.76	0.85	0.82	0.86	0.69	0.58	0.65	0.66	0.73	1.05
SOC	mean	0.1275	0.1436	0.1416	0.1322	0.1182	0.1189	0.1503	0.1405	0.1362	0.1161
	t-stat	2.71	2.56	2.53	2.25	2.21	2	2.49	2.36	2.48	3.36
	SR	0.82	0.81	0.78	0.71	0.66	0.64	0.82	0.74	0.75	1.04
GOV	mean	0.1240	0.1373	0.1375	0.1405	0.1380	0.1348	0.1445	0.1097	0.1232	0.1160
	t-stat	2.38	2.48	2.48	2.57	2.44	2.3	2.73	1.9	2.12	3.3
	SR	0.74	0.81	0.79	0.77	0.76	0.73	0.81	0.59	0.67	1.03

available yet in the database to give the trading signal for 2017.

Figure 4.3: Number of stocks traded in the percentile strategies

This figure shows the number of stocks included in the top and bottom percentile portfolios over the sample period from January 2003 to May 2017. The four portfolio cut-offs are represented by the dash-dotted (5%), dotted (10%), dashed (20%) and solid (30%) line. The upper line of the type denotes the full sample portfolio and the lower line denotes the portfolio with the negative screening.



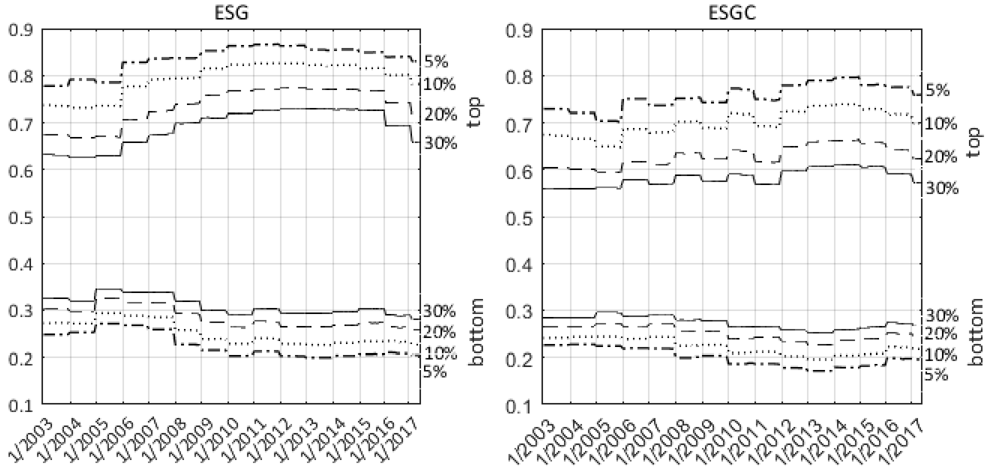
The following point of interest is the average score of the stocks traded in the particular percentiles. Figure 4.4 depicts the average score of the ESG and ESGC portfolios. In line with the thresholds of the decile portfolios, the score value of the top portfolios increases here over time as well. Noticeably, the average scores are not symmetrical for the top and bottom portfolios. Also, the bottom average scores of $5 - 30\%$ are not that spread out as the top $5 - 30\%$. At the end there is a slight drop in the scores of the top portfolios, which could mean that the stocks of the last index addition (Russel 2000) have generally of lower ratings.

Table 4.2 shows the correlations of the top and bottom portfolios of strategies considering different cut-offs. The long and short leg of the long-short strategy are highly correlated. The correlation increases as the size of the sample is bigger. the correlations of value-weighted portfolios range from 0.7917 to 0.9468. For example for value-weighted portfolio formed on the ESG scores, $Corr(ESG_{t5}, ESG_{b5}) = 0.8049$ while $Corr(ESG_{t30}, ESG_{b30}) = 0.9333$. For the equally weighted portfolios, the correlations are generally higher. Even for the 5% threshold the correlations are around 0.9 and above. For the 30% cut-off, the correlations reach as high as 0.98.

The high level of the correlations is in line with the exposure of these strategies to the market influences. Table 4.3 shows the correlations of the top and bottom strategies

Figure 4.4: Average scores of ESG and ESGC portfolios

The figure shows the average ESG and ESGC scores of different percentile portfolios over time. The upper (lower) part of the figure displays the average scores of portfolios including the highest (lowest) ranked stocks in the particular percentile. The solid, dashed, dotted and dashed-dotted lines denote the 5%, 10%, 20% and 30% portfolio cut-offs, respectively.



excess of the market return: $Corr(ESG_t - Rm, ESG_b - Rm)$. After removing the market return, the correlations of the top and bottom strategies lower substantially. This indicates that the long and short leg of the long-short strategy behave differently after the market element is accounted for. The right hand side of the table displays the correlation of the long-short strategies with the market portfolio. These are also considerably low and therefore the long-short strategies are considered suitable for further analysis as the influences of the broad market are small.

Finally, the overview of the top, bottom and long-short strategies follows. Table 4.4 shows the annualized mean excess returns and Sharpe ratios as performance evaluation measures. Overall, the mean returns are higher for the bottom strategies compared to the top. Therefore, the long-short strategy leads to negative mean excess returns. These are significantly different from zero for the 5 and 10% cut-off for the ESG, ESGC and ENV strategy. When comparing the strategies, the Sharpe ratio gives an idea of the relative performance of the strategies, however, this relation is not tested here as stated in the methodology section. For the ESG and ENV strategy, the Sharpe ratios are noticeably higher for the 5 and 10% cut-off cases compared to the other cut-offs. This suggests that such smaller subsamples produce more distinctive performance using the score signal than the large subsamples where the information from the score seems to be more diluted. In the ESGC strategy, only the 5% cut-off strategy gives seemingly higher performance than the larger subsamples. For SOC and GOV, no clear pattern appears.

When looking at the strict alternatives (strategies where companies from harmful

Table 4.2: Correlations of the top and bottom strategies

This table presents the correlations of the excess returns of the top and bottom strategies of the particular percentile and score type both for value- and equal-weighted portfolios.

	value weighted				equal weighted			
top/bottom	5%	10%	20%	30%	5%	10%	20%	30%
ESG	0.8049	0.8586	0.9147	0.9333	0.9174	0.9461	0.9671	0.9764
ESGC	0.8377	0.8935	0.9393	0.9434	0.9190	0.9527	0.9679	0.9746
ENV	0.8559	0.9030	0.9292	0.9453	0.9191	0.9485	0.9654	0.9781
SOC	0.8154	0.8744	0.9182	0.9347	0.9068	0.9436	0.9684	0.9808
GOV	0.8214	0.8549	0.8999	0.9345	0.9149	0.9432	0.9677	0.9804
strict ESG	0.7917	0.8612	0.9150	0.9334	0.9009	0.9341	0.9627	0.9739
strict ESGC	0.8148	0.8767	0.9318	0.9373	0.9179	0.9521	0.9599	0.9691
strict ENV	0.8499	0.8910	0.9245	0.9468	0.9167	0.9433	0.9603	0.9747
strict SOC	0.7971	0.8684	0.9232	0.9442	0.9005	0.9382	0.9674	0.9797
strict GOV	0.8191	0.8472	0.9090	0.9399	0.9063	0.9447	0.9659	0.9784

Table 4.3: Strategies and the correlation with the market

This table presents the correlations of the top and bottom strategies excess over the market. It also shows the correlations of the long-short strategies with the market portfolio. The long-short strategy holds long position in the top portfolio and short position in the bottom portfolio.

	Correlation of the top and bottom strategy excess over market				Correlation of the long-short strategy with the market			
cut-off	5%	10%	20%	30%	5%	10%	20%	30%
ESG	-0.2958	-0.4036	-0.5127	-0.5950	-0.0929	-0.2105	-0.3021	-0.3068
ESGC	-0.0107	0.1523	0.0407	-0.1334	-0.1107	0.0359	-0.0040	-0.0988
strict ESG	-0.1897	-0.1926	-0.2324	-0.2650	-0.0522	-0.0496	-0.2084	-0.2239
strict ESGC	-0.0264	0.1296	0.0713	0.0141	-0.0522	-0.0496	-0.2084	-0.2239

industries were divested) we can see that the mean returns of the strict ESG and ESGC long-short strategy are significantly different from zero even with 20% and for ESGC even for the 30% cut-off level. The other long-short strategies do not show mean returns significantly different from zero apart from the 20% SOC and 10% GOV strategy. For the strict ESG and ESGC the Sharpe ratios decline as we move from the smallest percentile to the larger ones. For the strict SOC and GOV strategy, Sharpe ratio stands out where the mean returns were significant (SOC_{ls20} and GOV_{ls10}).

The results for the equally weighted portfolios can be found in Table A.2 in Appendix A. The mean excess returns of these strategies are in general higher than the value-weighted portfolios. However, when the mean returns are adjusted for risk, the Sharpe ratios are of similar size and patterns.

Table 4.4: Value-weighted percentile strategies

This table presents the annualized mean excess returns and Sharpe ratios (SR) of the value-weighted percentile strategies. The top, bottom and long-short strategies for the cut-offs 5%, 10%, 20% and 30% are presented. The strict alternatives denote strategies with the negative screening. The values in the square brackets are the t-statistics from testing the mean returns against zero using the Newey-West standard errors.

cut-off		5%	10%	20%	30%	5%	10%	20%	30%
		ESG				strict ESG			
mean	top	0.0714	0.0793	0.0845	0.0903	0.0643	0.0760	0.0806	0.0873
		[1.60]	[2.00]	[2.07]	[2.17]	[1.39]	[1.80]	[1.89]	[2.03]
	bottom	0.1248	0.1262	0.1064	0.1039	0.1299	0.1150	0.1077	0.1028
		[2.37]	[2.55]	[2.07]	[2.06]	[2.59]	[2.62]	[2.20]	[2.07]
	L/S	-0.0480	-0.0421	-0.0200	-0.0123	-0.0586	-0.0353	-0.0247	-0.0141
		[-2.59]	[-2.64]	[-1.39]	[-0.96]	[-2.77]	[-2.50]	[-2.12]	[-1.20]
SR	top	0.53	0.62	0.65	0.68	0.48	0.56	0.61	0.65
	bottom	0.82	0.84	0.70	0.68	0.87	0.82	0.73	0.69
	L/S	-0.53	-0.55	-0.32	-0.22	-0.63	-0.48	-0.42	-0.26
		ESGC				strict ESGC			
mean	top	0.0873	0.0886	0.0941	0.0856	0.0878	0.0853	0.0906	0.0806
		[1.88]	[1.98]	[2.09]	[1.89]	[1.92]	[1.92]	[2.01]	[1.79]
	bottom	0.1669	0.1245	0.1140	0.1110	0.1602	0.1276	0.1144	0.1085
		[3.05]	[2.48]	[2.43]	[2.36]	[3.05]	[2.54]	[2.41]	[2.28]
	L/S	-0.0690	-0.0323	-0.0180	-0.0231	-0.0632	-0.0378	-0.0215	-0.0254
		[-3.35]	[-2.10]	[-1.94]	[-1.95]	[-2.62]	[-2.43]	[-1.98]	[-2.11]
SR	top	0.59	0.61	0.65	0.61	0.59	0.59	0.63	0.58
	bottom	1.00	0.83	0.78	0.76	0.96	0.85	0.78	0.74
	L/S	-0.76	-0.47	-0.36	-0.47	-0.65	-0.52	-0.40	-0.50
		ENV				strict ENV			
mean	top	0.0694	0.0820	0.0846	0.0878	0.0732	0.0795	0.0831	0.0852
		[1.59]	[1.92]	[2.01]	[2.15]	[1.66]	[1.83]	[1.89]	[2.00]
	bottom	0.1331	0.1294	0.1062	0.1041	0.1294	0.1256	0.1051	0.1061
		[2.17]	[2.37]	[1.94]	[1.96]	[2.00]	[2.26]	[1.90]	[2.01]
	L/S	-0.0568	-0.0424	-0.0197	-0.0149	-0.0502	-0.0413	-0.0201	-0.0191
		[-2.32]	[-2.43]	[-1.21]	[-1.01]	[-1.72]	[-1.92]	[-1.17]	[-1.35]
SR	top	0.48	0.59	0.63	0.66	0.49	0.56	0.60	0.63
	bottom	0.71	0.78	0.70	0.68	0.68	0.75	0.70	0.70
	L/S	-0.58	-0.59	-0.35	-0.29	-0.49	-0.54	-0.35	-0.39
		SOC				strict SOC			
mean	top	0.0772	0.0818	0.0819	0.0894	0.0750	0.0777	0.0784	0.0863
		[1.71]	[1.99]	[1.94]	[2.16]	[1.64]	[1.81]	[1.77]	[2.00]
	bottom	0.1183	0.1020	0.1096	0.1056	0.1152	0.1043	0.1119	0.1089
		[2.04]	[1.97]	[2.25]	[2.09]	[1.96]	[2.06]	[2.33]	[2.22]
	L/S	-0.0371	-0.0184	-0.0252	-0.0148	-0.0364	-0.0243	-0.0304	-0.0205
		[-1.59]	[-1.06]	[-1.98]	[-1.13]	[-1.50]	[-1.54]	[-2.54]	[-1.77]
SR	top	0.560	0.618	0.620	0.673	0.542	0.576	0.575	0.634
	bottom	0.675	0.638	0.724	0.687	0.659	0.663	0.763	0.720
	L/S	-0.365	-0.237	-0.417	-0.265	-0.345	-0.312	-0.539	-0.408
		GOV				strict GOV			
mean	top	0.1090	0.0895	0.0934	0.0929	0.1028	0.0756	0.0866	0.0890
		[2.87]	[2.07]	[2.26]	[2.20]	[2.59]	[1.73]	[2.02]	[2.02]
	bottom	0.1111	0.1167	0.1093	0.0948	0.1144	0.1219	0.1058	0.0918
		[1.89]	[2.22]	[2.09]	[1.90]	[2.34]	[2.51]	[2.10]	[1.88]
	L/S	-0.0019	-0.0245	-0.0145	-0.0018	-0.0105	-0.0417	-0.0176	-0.0026
		[-0.07]	[-1.10]	[-0.86]	[-0.13]	[-0.50]	[-2.03]	[-1.24]	[-0.23]
SR	top	0.80	0.67	0.74	0.72	0.74	0.56	0.67	0.67
	bottom	0.66	0.77	0.72	0.62	0.70	0.84	0.71	0.61
	L/S	-0.02	-0.31	-0.21	-0.03	-0.11	-0.54	-0.28	-0.05

4.2 Factor models

In this section the strategies will be evaluated by factor models. Here I will focus on the value-weighted strategies. Table 4.5 shows the results for the 10% cut-off ESG and ESGC strategies as well as the strict alternatives. The long-short ESG strategy has a monthly mean excess return of -0.36% which is statistically significant at the 1% level. After accounting for the market factor the alpha is -0.27% and for the 4-factor model the alpha increases to -0.25% with the significance going down to the 10% level for both models. When adding the 3 factors, adjusted R^2 increases from 0.04 to 0.17. The adjusted R^2 of 0.17 for the ESG 4-factor model and 0.11 for the strict ESG model is still considerably low. The relatively low explanatory power of the model is in line with the results in the literature. For example, Halbritter and Dorfleitner (2015) reports 4-factor model adjusted R^2 in the range from -0.004 to 0.44 for the long-short strategies and Statman and Glushkov (2009) present the 4-factor model with $R^2 = 0.19$. The loading on the SMB factor is statistically significant and negative, suggesting that the differences in returns of top and bottom ESG stocks have common characteristics as the differences in returns of big and small stocks. The loading on the HML factor is significantly positive, which indicates that the differences in returns of top and bottom ESG stocks have common characteristics as the differences in returns of high book-to-market and low book-to-market stocks. The performance of this long-short strategy comes mostly from the short leg where the monthly abnormal return is 0.21% while the return of the top strategy is not significantly different from zero. Very similar results apply to the strict ESG strategy.

The ESGC long-short strategy leads to an abnormal return of -0.28% significant at the 5% level. Here the market return in the one factor model does not explain the ESGC strategy returns and the factor model is insufficient as well with adjusted R^2 of only 0.02. The statistical and economic significance is the same for the 4-factor model as compared to the non-risk adjusted performance and therefore the Carhart (1997) model does not seem to help explain the abnormal returns. The same holds for the strict ESGC strategy.

Overall, the bottom and top strategies are highly determined by the market portfolio. The short positions on the bottom portfolios are the major (and negative) contributors to the performance of the long-short strategies. The Controversy overlay score leads to slightly higher abnormal returns than the plain strategies. Unlike the ESG strategies, the ESGC long-short strategy returns do not seem to carry any common characteristics as the size factor returns.

The results suggest that to aim for a positive abnormal return, the long position should be held on the bottom portfolios and the short position on the top portfolios.

Table 4.5: Factor model for the ESG, ESGC, strict ESG and strict ESGC value-weighted strategy

This table presents the benchmark models for the ESG, ESGC, strict ESG and strict ESGC strategies with the 10% cut-off. The long-short strategy holds a long position in the top-rated ESG stocks and short position in the bottom-rated ESG stocks. Models (1), (4) and (7) show the mean excess returns (above the risk-free rate) of the top, bottom and the long-short strategy, respectively. Models (2), (5) and (8) give the results of the one factor model with market return. Models (3), (6) and (9) present the results from the 4-factor model. All models are estimated over the sample period from January 2003 to May 2017 on a monthly basis. All portfolios are weighted by the firms' market capitalization. Adjusted R^2 is reported. The standard errors are estimated using the Newey and West (1987) method.

	(1)	(2)	(3)	(4)	(5)	(6) 7	(7)	(8)	(9)
	t10	t10	t10	b10	b10	b10	ls10	ls10	ls10
					ESG				
Constant	0.0064**	-0.0006	-0.0004	0.0100**	0.0021*	0.0021*	-0.0036***	-0.0027*	-0.0025*
	(0.0032)	(0.0006)	(0.0005)	(0.0039)	(0.0011)	(0.0011)	(0.0014)	(0.0014)	(0.0015)
Rm-rf		0.890***	0.926***		1.006***	0.992***		-0.116***	-0.066
		(0.015)	(0.026)		(0.032)	(0.045)		(0.040)	(0.067)
SMB			-0.238***			0.078			-0.317***
			(0.035)			(0.066)			(0.093)
HML			0.136***			-0.109*			0.245**
			(0.046)			(0.060)			(0.096)
WML			0.018			-0.041			0.059
			(0.014)			(0.042)			(0.054)
R^2		0.9247	0.9481		0.8461	0.8749		0.0387	0.1712
					ESGC				
Constant	0.0071**	-0.0007	-0.0008	0.0098**	0.0021*	0.0021*	-0.0027**	-0.0029**	-0.0028**
	(0.0036)	(0.0010)	(0.0010)	(0.0040)	(0.0012)	(0.0012)	(0.0013)	(0.0014)	(0.0014)
Rm-rf		1.001***	1.014***		0.983***	0.991***		0.017	0.023
		(0.041)	(0.038)		(0.025)	(0.027)		(0.040)	(0.039)
SMB			-0.025			0.046			-0.070
			(0.054)			(0.057)			(0.075)
HML			-0.060*			-0.226***			0.165***
			(0.033)			(0.066)			(0.057)
WML			-0.007			-0.055			0.048
			(0.021)			(0.042)			(0.031)
R^2		0.9174	0.9173		0.8597	0.8890		-0.0046	0.0202
					strict ESG				
Constant	0.0061*	-0.0011	-0.0008	0.0091***	0.0017*	0.0017*	-0.0030**	-0.0028**	-0.0025*
	(0.0034)	(0.0008)	(0.0007)	(0.0035)	(0.0009)	(0.0010)	(0.0012)	(0.0013)	(0.0014)
Rm-rf		0.919***	0.934***		0.945***	0.940***		-0.026	-0.006
		(0.025)	(0.031)		(0.031)	(0.042)		(0.045)	(0.060)
SMB			-0.224***			0.055			-0.280***
			(0.044)			(0.057)			(0.091)
HML			0.118			-0.062			0.181**
			(0.080)			(0.049)			(0.089)
WML			-0.035			-0.010			-0.026
			(0.026)			(0.031)			(0.039)
R^2		0.9029	0.9253		0.8424	0.8683		-0.0034	0.1109
					strict ESGC				
Constant	0.0069*	-0.0009	-0.0008	0.0101**	0.0024*	0.0024*	-0.0032**	-0.0033**	-0.0032**
	(0.0036)	(0.0011)	(0.0011)	(0.0040)	(0.0013)	(0.0013)	(0.0013)	(0.0014)	(0.0014)
Rm-rf		0.986***	0.997***		0.974***	0.980***		0.012	0.016
		(0.042)	(0.037)		(0.024)	(0.025)		(0.041)	(0.043)
SMB			-0.045			0.024			-0.070
			(0.056)			(0.061)			(0.079)
HML			-0.060			-0.206**			0.147**
			(0.055)			(0.0810)			(0.058)
WML			-0.024			-0.061*			0.037
			(0.026)			(0.037)			(0.031)
R^2		0.8974	0.8975		0.8539	0.8807		-0.0053	0.0062

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Such strategy would have produced monthly abnormal return of 0.25% for the ESG and strict ESG strategy and 0.28% and 0.32% for the ESGC and strict ESGC strategies, respectively.

The results for other cut-off levels and score types can be found in Table A.3 and Table A.4 in Appendix A. The ESG and strict ESG strategy shows higher significance levels considering the 5% cut-off level. However, the significance disappears with the 20% and 30% threshold. For the ESGC and strict ESGC strategy similar significance pattern appears but as described above, the 4-factor model has close-to-zero explanatory power. The ENV score portfolio has statistically significant (at 5% level) mean excess returns of −0.49% and −0.36% for the cut-off of 5% and 10%, respectively. However, the significance disappears when accounting for the market factor. The statistically significant mean returns in *strict* SOC_{ls20} lower with the market factor adjustment and disappear after accounting for other factors mainly due to a negative loading on SMB. For *strict* GOV_{ls10} the abnormal return is significant at 10% even after adjusting for risk with the 4-factor model. Nevertheless, the sustainability information does not seem to be strong enough within the environmental, social and governance pillars. On the other hand, the patterns with the ESG and ESGC strategies are clearer in the sense that the portfolios formed based on the most extreme scores exhibit the highest statistical significance which then lowers as the portfolio size gets bigger. When considering the equally weighted portfolio strategies, only the *strict* ESG_{ls5} and *strict* ESG_{ls10} show statistically significant abnormal returns. Generally, the R^2 of these models is higher than for the value-weighted portfolio strategies. This fact together with significant SMB loadings indicates that after accounting for the differences in return characteristics between the small and big firms, the returns of these equal-weighted strategies disappear.

For further analysis, I redefine the long-short strategies. The long position is now held in the bottom portfolios and the short position is held in the top portfolios. To give an overview of abnormal returns of all strategies, I present the annualized mean excess returns and annualized 4-factor alphas in Table 4.6.

These results (long on the bottom, short the top) show the 4-factor alpha of 4.25% per annum for the value-weighted ESG_{ls5} strategy and 6.6% per annum for $ESGC_{ls5}$. These are statistically and economically significant. These alphas decline to 3% and 3.4% for the ESG_{ls10} and $ESGC_{ls10}$ respectively, with only the ESGC alternative being significant at the 5% level. When lower thresholds are considered, the significance disappears. In general, the negative screening (strict strategies) do not harm the portfolio performance. The strict ESG_{ls5} and $ESGC_{ls10}$ strategies even produce 1.5 and 0.5 percentage points higher abnormal returns than their non-strict alternatives, respectively. The value-weighted portfolios are considered as a more suitable strategy here because

Table 4.6: Annualized mean excess returns and 4-factor alphas

This table presents the annualized mean excess returns and annualized 4-factor alphas of the long-short strategies estimated using the sample period from January 2003 to May 2017. These portfolios hold long position on stocks with the lowest scores and a short position on stocks with the highest scores. Portfolio cut-offs $5\%, 10\%, 20\%$ and 30% are used for the value weighted and equally weighted portfolios. The regressions are run for the full sample portfolios (not excluding stocks) and the strict alternatives (negative screening).

	value-weighted		equal-weighted		value-weighted		equal-weighted	
	mean	alpha	mean	alpha	mean	alpha	mean	alpha
	full sample strategies				strict strategies			
ESG_{ls5}	0.0501**	0.0425**	0.0408*	0.0282	0.0619***	0.0574**	0.0523**	0.0402**
ESG_{ls10}	0.0436***	0.0300*	0.0349**	0.0192*	0.0365**	0.0299*	0.0390**	0.0264**
ESG_{ls20}	0.0203	0.0051	0.0027	-0.0090	0.0252**	0.0135	0.0054	-0.004
ESG_{ls30}	0.0124	-0.0013	-0.0006	-0.0095	0.0143	0.0036	0.0024	-0.0063
$ESGC_{ls5}$	0.0736***	0.0659***	0.0319	0.0223	0.0671***	0.0608**	0.0335	0.0246
$ESGC_{ls10}$	0.0332**	0.0341**	0.0118	0.0062	0.0392**	0.0395**	0.0195	0.0136
$ESGC_{ls20}$	0.0183*	0.0177*	-0.0069	-0.0105	0.0219**	0.0211*	-0.0004	-0.0036
$ESGC_{ls30}$	0.0235*	0.0188	-0.0107	-0.0150	0.0260**	0.0210*	-0.0078	-0.0114
ENV_{ls5}	0.0599**	0.0344	0.0591**	0.0422*	0.0527*	0.0246	0.0651**	0.0467*
ENV_{ls10}	0.0441**	0.0296*	0.0347**	0.0256*	0.0429*	0.0261	0.0356**	0.0254*
ENV_{ls20}	0.0201	0.0105	0.0082	-0.0037	0.0205	0.0129	0.0146	0.0038
ENV_{ls30}	0.0151	0.0027	-0.0018	-0.0114	0.0194	0.0096	0.0015	-0.0068
SOC_{ls5}	0.0385	0.0177	0.0339	0.0179	0.0376	0.0168	0.0349	0.0185
SOC_{ls10}	0.0188	0.0003	0.0056	-0.0041	0.0249	0.0075	0.0094	0.0006
SOC_{ls20}	0.0257**	0.0127	-0.0001	-0.0076	0.0313**	0.0221	0.0074	0.0009
SOC_{ls30}	0.0150	0.0009	-0.0003	-0.0062	0.0210*	0.0097	0.0043	-0.0009
GOV_{ls5}	0.0019	-0.0183	0.0123	-0.0002	0.0106	-0.0059	0.0283	0.0161
GOV_{ls10}	0.0251	0.0141	0.0270	0.0164	0.0433*	0.0362*	0.0303	0.0212
GOV_{ls20}	0.0146	-0.0038	0.0071	-0.0045	0.0179	0.0027	0.0081	-0.0029
GOV_{ls30}	0.0018	-0.0150	-0.0034	-0.0136*	0.0026	-0.0118	-0.0034	-0.0126*

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the results from the equal-weighted portfolios seems to be driven by small stocks. However, the signal from the score information is not reflected in the returns when weighted equally. In the equal-weighted portfolios, the alphas are not statistically significant with the exception of the *strict* ESG_{ls5} and *strict* ESG_{ls10}. In both weighting schemes, utilizing the score information ENV, SOC, or GOV does not yield any significant risk-adjusted abnormal returns.

When we compare these results with studies using the approach of long position in the high score, short in the low score and adjusting with the 4-factor model, we can see that my results are of the opposite direction. That is since Statman and Glushkov (2009) finds annualized alpha of 5% and Kempf and Osthoff (2007) finds alpha of 4.9% when going long on the top-ranked stocks and short on the low-ranked stocks. These studies utilize the KLD database for scores in the sample period 1992-2004 (Kempf and Osthoff, 2007) and 1992-2007 (Statman and Glushkov, 2009). Later studies find no significance of the relation, for example Halbritter and Dorfleitner (2015) who employs the KLD, Bloomberg and ASSET4 databases during 1992-2012. Therefore, the only results showing positive statistically significant abnormal returns seem to be tied to the KLD database during time period until the middle of the 2000s' first decade. Based on this, I do not find my results contradictory to the research in the academic literature. These results are associated with the information from the scores given by Thomson Reuters ESG database during the specified sample period.

4.3 Sentiment hypothesis

In this section I am going to test whether the abnormal returns found by the 4-factor models could be explained by a sentiment-driven overpricing. Table 4.7 reports mean excess returns separately for the months when returns follow the high sentiment and for those following the low sentiment periods.

Table 4.7: Returns in periods following high and low sentiment

The table shows the mean excess returns in month following high and low sentiment periods, which are classified based on the median value of the Baker and Wurgler (2006) index adjusted for business cycle effects. The strategies included are the ESG and ESGC using the 5% and 10% cut-off as well as the strict strategies with negative screening. The long-short strategy holds a long position in the bottom portfolio (stocks with a low score) and short position in the top portfolio (stocks with a high score). The sample period is from January 2003 to October 2015 on a monthly basis. The values in square brackets are t-statistics computed with the Newey and West (1987) method.

	Bottom portfolios			Top portfolios			Long-short		
	High sentiment	Low sentiment	High -low	High sentiment	Low sentiment	High -low	High sentiment	Low sentiment	High -low
ESG_5	0.0064	0.0139	-0.0074	0.0029	0.0079	-0.0050	0.0035	0.0059	-0.0024
	[1.16]	[2.09]	[-1.18]	[0.84]	[1.24]	[-0.77]	[1.18]	[2.49]	[-0.54]
ESG_{10}	0.0084	0.0116	-0.0033	0.0046	0.0075	-0.0029	0.0038	0.0041	-0.0003
	[1.80]	[1.67]	[-0.45]	[1.61]	[1.19]	[-0.47]	[1.53]	[1.63]	[-0.09]
$ESGC_5$	0.0094	0.0156	-0.0062	0.0028	0.0102	-0.0074	0.0067	0.0054	0.0013
	[1.93]	[2.25]	[-0.77]	[0.64]	[1.56]	[-1.05]	[1.89]	[3.14]	[0.30]
$ESGC_{10}$	0.0070	0.0120	-0.0050	0.0038	0.0096	-0.0058	0.0032	0.0024	0.0008
	[1.46]	[1.81]	[-0.69]	[0.94]	[1.48]	[-0.83]	[1.41]	[1.40]	[0.25]
$strict\ ESG_5$	0.0070	0.0143	-0.0073	0.0021	0.0071	-0.0050	0.0049	0.0072	-0.0023
	[1.31]	[2.24]	[-1.02]	[0.58]	[1.04]	[-0.77]	[1.79]	[2.52]	[-0.05]
$strict\ ESG_{10}$	0.0089	0.0095	-0.0006	0.0041	0.0071	-0.0030	0.0048	0.0024	0.0024
	[2.07]	[1.47]	[-0.09]	[1.42]	[1.01]	[-0.46]	[2.19]	[1.26]	[0.69]
$strict\ ESGC_5$	0.0101	0.0132	-0.0031	0.0021	0.0103	-0.0082	0.0080	0.0029	0.0051
	[2.06]	[1.99]	[-0.39]	[0.51]	[1.54]	[-1.15]	[2.39]	[1.31]	[1.14]
$strict\ ESGC_{10}$	0.0071	0.0120	-0.0049	0.0030	0.0094	-0.0064	0.0040	0.0026	0.0014
	[1.44]	[1.85]	[-0.69]	[0.76]	[1.41]	[-0.91]	[1.84]	[1.60]	[0.42]

The results from this table are the first indicator whether the anomaly is more pronounced in the months following high sentiment period. In this case, the long-short strategies do not always exhibit higher mean excess returns in periods following high sentiment. Moreover, none of the differences between high and low sentiment periods is statistically significant. I will proceed with reporting only with the strategies ESG_{ls10} and $ESGC_{ls10}$ for further sentiment hypothesis testing as the results of the strategies with 5% cut-off and the strict strategies give the same implications.

Table 4.8: The 4-factor model with sentiment dummy variables

The table reports the abnormal returns following high and low sentiment periods for the strategies ESG and ESGC using 10% cut-off for the portfolio construction. The abnormal returns are the estimates of a_H and a_L from the model $R_t - R_{f,t} = a_H d_{H,t} + a_L d_{L,t} + b(Rm_t - R_{f,t}) + cSMB_t + dHML_t + eWML_t + \epsilon_t$, where $d_{H,t}$ and $d_{L,t}$ are dummy variables indicating whether the period follows high and low sentiment, respectively. The long-short strategy holds a long position in the bottom portfolio (stocks with a low score) and short position in the top portfolio (stocks with a high score). The sample period is from January 2003 to October 2015 on a monthly basis. The values in square brackets are t-statistics computed with the Newey and West (1987) method.

	Bottom portfolios			Top portfolios			Long-short		
	High sentiment	Low sentiment	High -low	High sentiment	Low sentiment	High -low	High sentiment	Low sentiment	High -low
ESG_{10}	0.0014	0.0018	-0.0003	0.0008	-0.0011	0.0019	0.0006	0.0029	-0.0023
	[0.76]	[0.93]	[0.58]	[0.76]	[-1.99]	[0.57]	[0.26]	[1.34]	[0.07]
$ESGC_{10}$	-0.0005	0.0023	-0.0028	-0.0015	-0.0001	-0.0014	0.0010	0.0024	-0.0014
	[-0.23]	[1.31]	[0.05]	[-0.80]	[-0.09]	[0.65]	[0.41]	[1.37]	[0.17]

Table 4.8 presents the results from 4-factor regressions using indicator variables to specify whether the period follows high or low sentiment. With this division of the sample, the abnormal return of the long-short strategies is no longer significant and neither is the difference between high and low sentiment alphas.

The final models with the sentiment incorporated are presented in Table 4.9. From the simple regression of the excess returns of the long-short ESG and ESGC strategies on the sentiment index we can see that the sentiment variable is statistically insignificant with close-to-zero adjusted R^2 of the model. Then the 4-factor model is re-estimated and after that the model with sentiment and factors is presented. When the lagged sentiment index is added to the model, it does not help to explain the excess returns. The coefficient of the sentiment variable is insignificant, the size and significance of alpha does not change and the adjusted R^2 decreases.

All in all, the relation of sentiment and anomalies of Stambaugh et al. (2012) has not been found in the socially responsible investing context. Table 4.7 and Table 4.8 already indicate that it cannot be shown that the anomaly is stronger in months following high sentiment periods. Table 4.9 is in line with this evidence, showing that incorporating the sentiment to the 4-factor model even lowers the explanatory power of the model.

To summarize this section, the results from these models do not provide enough evidence to claim that the abnormal returns of ESG and ESGC strategies are attributable

Table 4.9: Risk-adjusted returns and the sentiment index

Models (1) and (4) of the table report the regression results of the excess returns of ESG and ESGC long-short strategy (using the 10% portfolio cut-off) on the lagged value of the level of the sentiment index (BW) of Baker and Wurgler (2006) adjusted for business cycle effects. Models (2) and (5) estimate the 4-factor model of Carhart (1997) and models (3) and (6) add the lagged value of the sentiment index. The long-short strategy holds a long position in the bottom portfolio (stocks with a low score) and short position in the top portfolio (stocks with a high score). The adjusted R^2 is reported. The sample period is from January 2003 to October 2015 on a monthly basis. The standard errors are computed using the Newey and West (1987) method.

	(1) ESG_{ls10}	(2) ESG_{ls10}	(3) ESG_{ls10}	(4) $ESGC_{ls10}$	(5) $ESGC_{ls10}$	(6) $ESGC_{ls10}$
Constant	0.0040***	0.0032**	0.0032**	0.0028**	0.0029**	0.0029**
	(0.0013)	(0.0015)	(0.0016)	(0.0013)	(0.0012)	(0.0013)
BW	-0.0021		-0.0006	-0.0012		-0.0010
	(0.0012)		(0.0015)	(0.0013)		(0.0013)
Rm-rf		0.0426	0.0427		-0.0263	-0.0261
		(0.0639)	(0.0640)		(0.0423)	(0.0421)
SMB		0.376***	0.372***		0.0816	0.0744
		(0.109)	(0.105)		(0.0849)	(0.0832)
HML		-0.268**	-0.270**		-0.128*	-0.131*
		(0.110)	(0.110)		(0.0736)	(0.0729)
WML		-0.0450	-0.0424		-0.0404	-0.0360
		(0.0517)	(0.0498)		(0.0303)	(0.0295)
R^2	0.0048	0.1845	0.1795	-0.0031	-0.0002	-0.0045

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

to a sentiment-driven mispricing. Instead, the abnormal returns might arise as a compensation for hidden risk factors. Low ESG/ESGC score possibly reflects risk and therefore these stocks are traded with a premium for such non-sustainability risk. However, is the significant alpha a sign of 'pure' ESG risk or does it reflect some other risks? This could be further tested by the regressions of Fama and MacBeth (1973) and then incorporating the zero-net investment factor into the model. This way, the pure signal from ESG would be extracted and an independent source of risk would be obtained. Such phenomenon would then be the ESG anomaly in the cross-section of returns. Then by integrating this factor into the model the portfolio could be more appropriately risk-adjusted. However, this is outside the scope of this thesis.

The implications of the results are that environmentally and socially-conscious investors would experience severe losses by holding a long position on the high ranked stocks and a short position on the bottom ranked stocks. Excluding sin companies from the sample does not change these implications. Therefore, by utilizing the information from the Thomson Reuters ESG database, socially responsible investing does not relate to positive abnormal returns, instead the investors would experience a possibly severe harm to their portfolio. The values-driven investors should therefore rather stick to the impact investing realm and invest into projects that are in line with their personal beliefs. On the other hand, responsible investing could be attractive for purely profit-driven investors who could make use of the relation through active stock picking and using the ESG score to hold the bottom-ranked stocks long and the top-ranked stocks short. Nevertheless, such outcome would undermine one of the main purposes of compiling these databases, which is encouraging companies towards more responsible behavior.

These results are quite against the initial expectations, which can be shown by answering the main question: Is socially responsible investing attractive only for the environmentally/socially conscious investors or could it also attract purely profit-driven investors? Here I get back to the difference between the original definition of SRI and Responsible Investing. SRI is not attractive even for the values-driven investors because it causes severe harm to the portfolio performance and these investors have other investment opportunities to follow their values. Neither is it suitable for profit-driven investors as it requires long positions in the high ESG stocks. Responsible Investing could be the attractive approach for profit-driven investors where they could construct trading strategies with utilizing the ESG information in the opposite direction.

4.4 Robustness tests

The 4-factor model is re-estimated for all long-short strategies using the adjusted time span ending in December 2015. This step is justified by the considerably changing characteristics of the sample in 2016 as shown in the Data section.

Table 4.10 presents the results, which are in line with the previous factor model for the full sample period. The previously significant strategies remain statistically significant and in many cases the significance is of a higher level with a higher abnormal return. The annualized alpha of value-weighted ESG_{ls5} increases from 4.25% to 4.8%. The $ESGC_{ls5}$ strategy produces a noteworthy alpha of 6.74% per annum significant at 1% level. The ESGC strategy remains significant all the way with shifting the sample cut-off until 30%, which holds also for the strict ESGC strategy. Overall, these results suggest that the ESG and ESGC strategies seem to be even more profitable when the final part of the sample period, where the signals from the scores are deemed to be of lower quality, is removed from the analysis.

Next, the four-factor model is re-estimated using two subperiods: January 2003 - June 2009 ans July 2009 to December 2015. Table 4.11 presents the annualized mean excess returns and annualized alphas from the 4-factor model. We can see that the performance

Table 4.10: Annualized mean returns and 4-factor alphas using shorter sample period

This table presents the annualized mean excess returns and 4-factor alphas of the long-short strategies over the sample period January 2003 to December 2015. These portfolios hold long position on stocks with the lowest scores and a short position on stocks with the best scores. Portfolio cut-offs 5%, 10%, 20% and 30% are used for the value weighted and equally weighted portfolios. The regressions are run for the full sample portfolios (not excluding stocks) and the strict alternatives (negative screening).

	value-weighted		equal-weighted		value-weighted		equal-weighted	
	mean	alpha	mean	alpha	mean	alpha	mean	alpha
	full sample strategies				strict strategies			
ESG_{ls5}	0.0557***	0.0482**	0.0493*	0.0464***	0.0723***	0.0684***	0.0656**	0.0625***
ESG_{ls10}	0.0468***	0.0320*	0.0377**	0.0299***	0.0428***	0.0341*	0.0435**	0.0341**
ESG_{ls20}	0.0181	0.0078	0.0032	0.0041	0.0235**	0.0157	0.0061	0.0181
ESG_{ls30}	0.0096	0.0005	0.0006	0.0074	0.0110	0.0035	0.0018	0.0604
$ESGC_{ls5}$	0.075***	0.0674***	0.0326	0.0289	0.0677**	0.0622**	0.0372	0.0328
$ESGC_{ls10}$	0.0334**	0.0332**	0.0141	0.0086	0.0411**	0.0402**	0.0218	0.0169
$ESGC_{ls20}$	0.0186**	0.0188**	0.0058	0.0104	0.0224**	0.0222**	0.0000	0.0058
$ESGC_{ls30}$	0.0267**	0.0234*	0.0084	0.0132	0.0284**	0.0251**	0.0072**	0.0132
ENV_{ls5}	0.0611**	0.0400	0.0573*	0.0496**	0.0549	0.0310	0.0647**	0.0528**
ENV_{ls10}	0.0469**	0.0351*	0.0348*	0.0327**	0.0469*	0.0326	0.0364*	0.0291**
ENV_{ls20}	0.0176	0.0109	0.0069	0.0022	0.0196	0.0146	0.0134	0.0039
ENV_{ls30}	0.0100	0.0009	0.0038	0.0118	0.0146	0.0077	0.0003	0.0079
SOC_{ls5}	0.0380	0.0225	0.0312	0.0247	0.0387	0.0229	0.0317	0.0230
SOC_{ls10}	0.0184	0.0394	0.0000	0.0056	0.0261	0.0124	0.0042	0.0017
SOC_{ls20}	0.0252**	0.0161	0.0032	0.0094	0.0315**	0.0260*	0.0035	0.0024
SOC_{ls30}	0.0144	0.0044	0.0009	0.0054	0.0205*	0.0136	0.0027	0.0013
GOV_{ls5}	0.0032	0.0123	0.0201	0.0127	0.0139	0.0021	0.0414*	0.0365
GOV_{ls10}	0.0374	0.0299	0.0272**	0.0412***	0.0575***	0.0534**	0.0443***	0.0415***
GOV_{ls20}	0.0164	0.0017	0.0129	0.0055	0.0199	0.0077	0.0132	0.0039
GOV_{ls30}	0.0002	0.0136	0.0025	0.0088	0.0015	0.0104	0.0034	0.0108

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

of the 2003-2015 sample strategy is mainly driven by the later subperiod. The ESG_{ls10} strategy yields annualized alpha of 7.13% significant at 5% level in the second period as compared to only at the 10% level significant 3.2% in the 2003-2015 sample. The ESGC strategy also outperforms in the second sample period.

Finally, Table 4.12 presents the results from the five-factor model and demonstrates the change in the factor loadings beween the four- and the five-factor model. All factor models including the profitability and investment factors exhibit statistically significant alphas. In fact, the risk-adjusted returns are generally higher than in the 4-factor model and the investment factor CMA has a significant negative loading. The adjusted R^2 increases when adding these additional factors, however, in case of the ESGC strategy is still close to zero. Overall, the differences in returns of companies with low and high scores seem to have common characteristics with the differences in returns of aggressively and conservatively investing companies. The annualized alpha of the ESG_{ls10} strategy increases from 3.4% in the 4-factor model to 5.5% in the five-factor model. For the $ESGC_{ls10}$ the annualized abnormal return is now 4.9%.

To summarize the results from the checks performed by adjusting the sample or

Table 4.11: Annualized mean excess returns and 4-factor alphas during subperiods

This table presents the annualized mean excess returns and annualized 4-factor alphas of the long-short strategies over the sample period January 2003 to December 2015 as well as subperiods January 2003 to June 2009 and July 2009 to December 2015. These portfolios hold long position on stocks with the lowest scores and a short position on stocks with the highest scores. Portfolio cut-off of 10% is used for market capitalization weighted portfolios. The regressions are run for the full sample portfolios (not excluding stocks) and the strict alternatives (negative screening).

	01/2003 - 12/2015		01/2003 - 06/2009		07/2009 - 12/2015	
	mean	alpha	mean	alpha	mean	alpha
ESG_{ls10}	0.0468***	0.0320*	0.0235	0.0149	0.0706**	0.0713**
$ESGC_{ls10}$	0.0334**	0.0332**	0.0138	0.0072	0.0534**	0.0575**
strict ESG_{ls10}	0.0428***	0.0341*	0.0361	0.0339	0.0494**	0.0538*
strict $ESGC_{ls10}$	0.0411**	0.0402**	0.0258	0.0218	0.0566**	0.0604**

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

the theoretical model, the abnormal returns of the ESG and ESGC strategies (both full sample and strict) do not disappear in the long-run. The strategies exhibit better performance when the latest part of the sample period is excluded from he analysis. The long-term strategies seem to draw the abnormal returns mainly from the later part of the sample. Finally, the abnormal returns do not fade away when controlling for the exposure to the Fama and French (2015) quality factors. On the contrary, this model seems to better explain the style returns and also results in higher abnormal returns.

Table 4.12: Five-factor model

Models (1), (3), (5) and (7) of the table report the regression results of the Carhart (1997) 4-factor model. Models (2), (4), (6) and (8) estimate the 5-factor model of Fama and French (2015). Results include the ESG and ESGC long-short strategy (both full sample and with negative screening) using the 10% portfolio cut-off. The long-short strategy holds a long position in the bottom portfolio (stocks with a low score) and short position in the top portfolio (stocks with a high score). The adjusted R^2 is reported. The sample period is from January 2003 to December 2015 on a monthly basis. The standard errors are computed using the Newey and West (1987) method.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ESG_{ls10}		$ESGC_{ls10}$		$strict\ ESG_{ls10}$		$strict\ ESGC_{ls10}$	
Alpha	0.0026*	0.0049***	0.0027**	0.004**	0.0028*	0.0045**	0.0033**	0.0045**
	(0.0013)	(0.0016)	(0.0014)	(0.0017)	(0.0015)	(0.002)	(0.0014)	(0.0019)
Rm-rf	0.0743	0.0306	-0.0151	-0.0515	0.0122	-0.0251	-0.0086	-0.0560
	(0.0706)	(0.0369)	(0.0425)	(0.0552)	(0.0650)	(0.0395)	(0.0456)	(0.0629)
SMB	0.332***	0.268***	0.0808	0.0241	0.287***	0.261***	0.0787	0.0851
	(0.102)	(0.0955)	(0.0814)	(0.100)	(0.0978)	(0.0950)	(0.0848)	(0.102)
HML	-0.280***	-0.242*	-0.136**	-0.0717	-0.189*	-0.246**	-0.118*	-0.0254
	(0.106)	(0.135)	(0.0582)	(0.132)	(0.106)	(0.118)	(0.0615)	(0.124)
WML	-0.0619		-0.0471		0.0311		-0.0359	
	(0.0553)		(0.0308)		(0.0403)		(0.0297)	
RMW		-0.381*		-0.222		-0.265		-0.200
		(0.201)		(0.225)		(0.200)		(0.221)
CMA		-0.682***		-0.312**		-0.430**		-0.401***
		(0.126)		(0.130)		(0.182)		(0.129)
R^2	0.1838	0.2569	0.0065	0.0141	0.1159	0.1326	-0.0047	0.0292

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Chapter 5

Discussion

The quality of the research in the area of Socially Responsible Investing is subject to the quality of the database that provides the ESG information. It is therefore important to point out that also the results of any SRI empirical analysis are tied to the information extracted from the particular ESG scores database. The same holds for this thesis and therefore I would like to outline some disadvantages of the scores used for this analysis and areas for future improvement. As Dorfleitner et al. (2015) state, the properties of the ESG scores distribution largely differ considering different database providers. Therefore, the results of this thesis are not viewed as the resolution of the relation between SRI and portfolio performance in general, but more as an outcome dependent on the Thomson Reuters ESG database.

One of the obvious properties of this database is that the characteristics of an average company change over time as new indexes are included into the database. The most pronounced change could be seen in 2016 when the Russel 2000 index was being included into the database. Since it is an index of small-cap companies, the effect was considerable as could be seen in Figure 3.2. Large amount of companies were added and therefore the year 2016 represents a milestone and makes it challenging to interpret any empirical analysis including this period. For future empirical studies, the division of sample to pre-2016 and post January 2016 would be advisable.

Another issues is that the percentile ranking methodology orders the companies according to their performance, however, disregards the size of the performance differences among the companies. Based on an alternative rating, the portfolio weights could reflect also the level of the company ESG. The future enhancements of the database could also include controversy score for the pillars or even categories. Right now we can see only an overall Controversy score, however, it would be useful to see which area the controversy came from so that the controversy overlap could be done with ENV, SOC, and GOV scores separately.

What could be considered a bias in the database is that if the criteria cannot be found in publicly disclosed information, the company gets zero, which lowers the final ESG. Then the scores favor large-cap stocks because large companies may be able to disclose the information in a better way (have more resources to do so) and then it is easier for the ranking agency to find this information. This is often referred to as the problem of 'doing good' versus 'reporting good'. The responsibility of small companies may not be reflected in the scores while the disguised responsibility might be manipulated into the score by the big companies. This issue makes SRI less appealing since it is based on 'reporting good' and not on 'doing good' as is the case of impact investments.

Another critique of ESG reporting is that the level of the score does not necessarily reflect the attractiveness of the company for the environmentally and/or socially conscious investors. More specifically, the ESG score is not fully connected with the impact of the company's products. Tesla Inc. is an example of a company with a positive effect of its product but with low levels of the ESG scores (ESG of 0.28 and ESGC of 0.19 in 2016).

Next, we should consider the timing of the ESG score publication. As could be seen during the analysis, not all the scores for 2016 were available yet in the database to give the trading signal. Therefore, another robustness test would be desirable to make sure that the signal information was known at the time of the portfolio formation. The analysis could be redone considering different lags in months. Even though this approach is not considered in the academic literature, this aspect is important for the credibility and subsequent implementation of such investment strategies.

The results of this thesis and their interpretation need to be associated with the motivations of the investors. Even though there is no consensus of whether the SRI should be connected with the negative screening, the advocates of this investing approach argue that it an integral part of the investment process. Since the investors choose SRI because of their personal values, they would certainly also conduct the negative screening to avoid industries that are incompatible with their values. In case of the ESG database used in this thesis, these investors would also utilize the ESGC overlay score to account for the company controversies. They would therefore most likely choose the *strict ESGC* strategy. On the other hand, the profit-driven investors would not exclude sin companies because of diversification issues. Then they would likely choose the ESGC strategy as it exhibits better performance than the ESG strategy. Interestingly, the two groups of investors would thus end up holding the opposite positions (long vs. short) on the same exact stocks.

Lastly, it is crucial to comment on the profitability of the strategies. The abnormal returns obtained through this analysis do not account for any transaction costs. Therefore, it is unclear whether these strategies would be profitable after incorporating the transaction costs. These are usually associated with the trading volume. This is particularly interesting because of the significant abnormal returns found in the second half of the sample period. In this period, the average volume traded was approximately twice as large as the average volume traded in the first half of the sample (see Figure 4.3). Then the question arises whether the abnormal returns in the second subperiod would be washed away after transaction costs were incorporated into the analysis.

Chapter 6

Conclusion

This thesis offers a thorough insight into the Socially Responsible Investing realm. It investigates, whether investing into companies based on their CSR is a suitable investment approach for values-driven and profit-driven investors. It is an area of investments that enables active management by choosing stocks based on the company responsibility measured by the ESG characteristics. Empirical study in this area is then of a high interest because certain investors are seeking investment opportunities that are in line with their personal values. Then the question arises, whether these environmentally/socially conscious investors would have to give up good financial performance of their portfolio in return for investing in companies that are in line with their personal convictions. The goal of the thesis was to evaluate the portfolio performance of trading strategies using information on the companies' CSR and to look at how incorporating sentiment into the analysis could help explain the results.

The empirical analysis was based on the Thomson Reuters ESG Scores database provided by Thomson Reuters solely for the purpose of this thesis. A detailed description of the database highlighted the changing characteristics of the sample and also the bias towards large companies. The ESGC score, which overlays the pure ESG score with company controversies, was found to be one of the perks of the database. I extracted the U.S. sample from the database and showed that the sample has lower ESG levels than the global one, especially in the environmental sphere. Companies operating in harmful industries were found to have comparable levels of ESG as the no-harm alternatives. I used both active and dead companies for the analysis to avoid survivorship bias. The data used for the analysis was appropriately screened both in the static and time-series manner.

Portfolios were constructed during time period from January 2003 to May 2017 based on the ESG signal and also the controversy overlay, environmental, social and governance ratings were utilized. Zero investment strategies were designed by holding a long position in the top 5%, 10%, 20% or 30% of the sample and short position in the corresponding bottom counterpart according to the score rating. Strict strategies were also built using the negative industry screening. All strategies were evaluated first by Sharpe ratios and then by the four-factor model including the market, size, value and momentum factors.

The results of the empirical analysis show that long-short strategy holding a long position in the top ESG-ranked stocks and short position in the bottom stocks formed with the market capitalization weights leads to statistically and economically significant negative mean excess returns when extracting the ESG information from the outer 10% and 20% of the score distribution. When adjusted for common risk factors, strategy designed in a reversed manner (long on bottom, short on top) shows annualized abnormal return of 3.4% by employing the ESGC strategy with the 10% cut-off rate and the alternative with the negative screening reaches abnormal return of 3.95% per annum before transaction costs. The performance is shown to come mostly from the short leg of these strategies. Both the size and the statistical significance disappears when moving to the 20% and 30% cut-off strategies. Environmental, social and governance portfolio strategies do not exhibit any statistically significant abnormal returns.

The abnormal returns found cannot be explained by sentiment-driven mispricing. Therefore it is believed that the ESG score might represents additional source of risk. Robustness results suggest that the abnormal returns come mostly from the second part of the sample period. The model with the adjustment for investment and profitability factors seem to have higher explanatory power for these returns and the alphas even increase.

The implications of these results are twofold. First, SRI does not represent an attractive investment strategy for the value-minded investors using the Thomson Reuters ESG score database. It is advisable that they fully diversify their portfolio and then make use of the returns obtained to choose projects to invest in that are in line with their values. Second, purely profit-seeking investors could set up strategies based on the ESG rankings.

However, it remains unanswered to what degree the ESG score contributes to the risk-return profile of the investment strategy. Ideally, the ESG scores would be used to construct an investment factor, which could help to resolve this matter. Another limitation of this study is the missing lag structure of the portfolio formations to ensure the score availability. Such considerations would help with the practical implementation of the strategies and provide additional robustness check of the results. Finally, examining the effect of negative screening is subject to the sin stock definition in this analysis. However, negative screening remains a challenge as there is no proper definition of sin stocks and it is also likely to depend on the investor's personal decisions.

In future research using the Thomson Reuters ESG database it would also be desirable to evaluate trading strategies based on the scores from particular categories. This study used only the information aggregated into environmental, social and governance pillar, however, some categories could be found more suitable to give the information for the trading signal. Additionally, an alternative portfolio weighting could be used. Stocks that do not comply with the ESG level criteria would not be excluded from the portfolio, but rather they would be assigned lower weights. Also, another strategy could be constructed where the signal for trading would not be based on the level of ESG, but could utilize the information of the change in the ESG score. Then the strategy would invest in companies depending on whether their ESG ranking has increased or decreased as compared to the last score available.

An alternative approach for the value-minded investor would be to modify the classical portfolio selection. Instead of the mean-variance analysis, the ESG score could be incorporated into the model as the third dimension and play role in portfolio decision making. Then the investor would optimize given his/her appetite for risk, return and ESG responsibility of the portfolio.

Bibliography

Adler, T. and Kritzman, M. (2008). The cost of socially responsible investing. *The Journal of Portfolio Management*, 35(1):52–56.

Auer, B. R. and Schuhmacher, F. (2016). Do socially (ir) responsible investments pay? new evidence from international esg data. *The Quarterly Review of Economics and Finance*, 59:51–62.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.

Bebchuk, L. A., Cohen, A., and Wang, C. C. (2013). Learning and the disappearing association between governance and returns. *Journal of Financial Economics*, 108(2):323–348.

Blitz, D. and Fabozzi, F. J. (2017). Sin stocks revisited: Resolving the sin stock anomaly. *The Journal of Portfolio Management*, 44(1):105–111.

Bodie, Z., Kane, A., and Marcus, A. (2014). *Investments*. McGraw-Hill, New York.

Brammer, S., Brooks, C., and Pavelin, S. (2006). Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial management*, 35(3):97–116.

Cai, Y., Jo, H., and Pan, C. (2012). Doing well while doing bad? CSR in controversial industry sectors. *Journal of Business Ethics*, 108(4):467–480.

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

Chasan, E. (2017). What millennials want from their investments. Sustainable Finance. Bloomberg L.P., May 4, 2017.

Colby, L. (2017). Global sustainable investments grow 25% to $23 trillion. Bloomberg, L.P., March 27, 2017.

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

Dorfleitner, G., Halbritter, G., and Nguyen, M. (2015). Measuring the level and risk of corporate responsibility–an empirical comparison of different ESG rating approaches. *Journal of Asset Management*, 16(7):450–466.

Fabozzi, F. J., Ma, K., and Oliphant, B. J. (2008). Sin stock returns. *The Journal of Portfolio Management*, 35(1):82–94.

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

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1 – 22.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Ferrell, A., Liang, H., and Renneboog, L. (2016). Socially responsible firms. *Journal of Financial Economics*, 122(3):585–606.

Friedman, M. (September 13, 1970). The social responsibility of business is to increase its profits. *New York Times Magazine*, pages 173–178.

Guenster, N., Bauer, R., Derwall, J., and Koedijk, K. (2011). The economic value of corporate eco-efficiency. *European Financial Management*, 17(4):679–704.

Halbritter, G. and Dorfleitner, G. (2015). The wages of social responsibility—where are they? a critical review of ESG investing. *Review of Financial Economics*, 26:25–35.

Harvey, C. R. and Liu, Y. (2014). Evaluating trading strategies. *The Journal of Portfolio Management*, 40(5):108–118.

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

Hebb, T. (2013). Impact investing and responsible investing: what does it mean? *Journal of Sustainable Finance & Investment*, 3(2):71–74.

Höchstädter, A. K. and Scheck, B. (2015). What's in a name: An analysis of impact investing understandings by academics and practitioners. *Journal of Business Ethics*, 132(2):449–475.

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

Ince, O. S. and Porter, R. B. (2006). Individual equity return data from Thomson Datastream: Handle with care! *Journal of Financial Research*, 29(4):463–479.

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

Kempf, A. and Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, 13(5):908–922.

Kim, I. and Venkatachalam, M. (2011). Are sin stocks paying the price for accounting sins? *Journal of Accounting, Auditing & Finance*, 26(2):415–442.

Kim, Y., Li, H., and Li, S. (2014). Corporate social responsibility and stock price crash risk. *Journal of Banking & Finance*, 43:1–13.

Liang, H. and Renneboog, L. (2017). On the foundations of corporate social responsibility. *The Journal of Finance*, 72(2):853–910.

Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.

Mattingly, J. E. and Berman, S. L. (2006). Measurement of corporate social action discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1):20–46.

Mendell, M. and Barbosa, E. (2013). Impact investing: a preliminary analysis of emergent primary and secondary exchange platforms. *Journal of Sustainable Finance & Investment*, 3(2):111–123.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Pardini, D. (2017). ING Dives Into Loans That Link Interest Rates to ESG. Bloomberg L.P., Sustainable Finance, May 4, 2017.

PRI (2017). About the Principles of Responsible Investment. Principles of Responsible Investment. Retrieved on May 13, 2017 from https://www.unpri.org/about.

Revill, J. (2017). An insurance giant is shifting $130 billion into ethical investments. Retrieved on October 3, 2017 from http://www.businessinsider.com/r-swiss-re-shifts-130-billion-investments-to-track-ethical-indices-2017-7.

Rowley, T. and Berman, S. (2000). A brand new brand of corporate social performance. *Business & society*, 39(4):397–418.

Sibley, S. E., Wang, Y., Xing, Y., and Zhang, X. (2016). The information content of the sentiment index. *Journal of Banking & Finance*, 62:164–179.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Statman, M. and Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4):33–46.

Thomson Reuters (2017). Thomson Reuters ESG Scores. Retrieved on May 10, 2017 at https://financial.thomsonreuters.com/content/dam/openweb/documents/pdf/financial/esg-scores-methodology.pdf.

Ullmann, A. A. (1985). Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of us firms. *Academy of Management Review*, 10(3):540–557.

Appendix A

Tables of results

Table A.1: Decile strategies for portfolios with negative screening

This table reports the annualized mean excess returns and Sharpe ratios (SR) of the decile portfolios formed with value weights and equal weights. The t-statistic comes from testing the excess return series for a null hypothesis of zero mean using the Newey-West standard errors. Decile 1 denotes the top decile (portfolio formed from the stocks with the best scores) and decile 10 denotes the bottom portfolio (stocks with the lowest scores).

	decile	1	2	3	4	5	6	7	8	9	10
						value-weighted					
strict ESG	mean	0.0760	0.0945	0.1166	0.1100	0.0940	0.0879	0.0985	0.0969	0.0951	0.0881
	t-stat	1.80	2.10	2.38	2.24	1.74	1.72	1.80	1.90	1.69	2.53
	SR	0.56	0.68	0.75	0.70	0.57	0.54	0.61	0.60	0.58	0.84
strict ESGC	mean	0.0853	0.0955	0.0756	0.1059	0.0536	0.1108	0.1164	0.0966	0.1031	0.0888
	t-stat	1.92	2.04	1.62	2.55	1.11	2.30	2.24	1.82	2.21	2.55
	SR	0.59	0.63	0.53	0.80	0.36	0.70	0.74	0.60	0.66	0.83
strict ENV	mean	0.0795	0.0928	0.0959	0.1268	0.0966	0.0817	0.0773	0.1064	0.0969	0.0901
	t-stat	1.83	2.00	2.47	2.61	1.93	1.68	1.49	2.07	1.66	2.48
	SR	0.56	0.66	0.70	0.79	0.62	0.53	0.46	0.67	0.64	0.84
strict SOC	mean	0.0777	0.0812	0.1190	0.0968	0.1031	0.0898	0.1075	0.1025	0.1192	0.0853
	t-stat	1.81	1.67	2.86	2.02	2.12	1.70	1.91	1.90	2.57	2.41
	SR	0.58	0.56	0.81	0.64	0.66	0.59	0.65	0.61	0.81	0.80
strict GOV	mean	0.0756	0.1044	0.0989	0.0876	0.1189	0.1048	0.1025	0.0730	0.0926	0.0890
	t-stat	1.73	2.39	2.02	1.95	2.24	2.25	1.87	1.47	1.65	2.55
	SR	0.56	0.79	0.66	0.59	0.75	0.66	0.61	0.45	0.56	0.85
						equal-weighted					
strict ESG	mean	0.1166	0.1425	0.1374	0.1602	0.1186	0.1290	0.1257	0.1334	0.1116	0.1135
	t-stat	2.40	2.66	2.42	2.77	1.95	2.32	2.04	2.27	1.89	3.28
	SR	0.75	0.86	0.74	0.85	0.62	0.69	0.67	0.70	0.63	1.01
strict ESGC	mean	0.1322	0.1463	0.1340	0.1299	0.1162	0.1599	0.1318	0.1074	0.1237	0.1138
	t-stat	2.49	2.67	2.32	2.38	1.98	2.70	2.29	1.86	2.11	3.27
	SR	0.77	0.76	0.76	0.74	0.64	0.86	0.74	0.61	0.67	1.01
strict ENV	mean	0.1269	0.1325	0.1402	0.1562	0.1338	0.1018	0.1255	0.1250	0.1269	0.1149
	t-stat	2.35	2.60	2.59	2.59	2.43	1.76	2.17	2.16	2.00	3.31
	SR	0.74	0.82	0.80	0.80	0.74	0.56	0.66	0.67	0.68	1.01
strict SOC	mean	0.1255	0.1357	0.1459	0.1307	0.1189	0.1103	0.1465	0.1436	0.1417	0.1133
	t-stat	2.62	2.38	2.63	2.19	2.20	1.82	2.40	2.46	2.49	3.21
	SR	0.80	0.75	0.80	0.70	0.65	0.59	0.81	0.75	0.78	1.01
strict GOV	mean	0.1176	0.1417	0.1372	0.1357	0.1483	0.1283	0.1403	0.1087	0.1259	0.1126
	t-stat	2.29	2.56	2.48	2.45	2.48	2.15	2.60	1.81	2.12	3.22
	SR	0.70	0.84	0.77	0.73	0.80	0.69	0.78	0.58	0.69	1.00

Table A.2: Equal-weighted percentile strategies

This table presents the annualized mean excess returns and Sharpe ratios (SR) of the equal-weighted percentile strategies. The top, bottom and long-short strategies for the cut-offs $5, 10, 20$ and 30% are presented. The strict alternatives denote strategies with the negative screening. The values in the square brackets are the t-statistics from testing the mean returns against zero using the Newey-West standard errors.

		5%	10%	20%	30%	5%	10%	20%	30%
		ESG				strict ESG			
mean	top	0.1105	0.1210	0.1314	0.1358	0.1048	0.1166	0.1295	0.1321
		[2.12]	[2.49]	[2.62]	[2.60]	[2.02]	[2.40]	[2.55]	[2.53]
	bottom	0.1554	0.1597	0.1344	0.1352	0.1621	0.1597	0.1356	0.1348
		[2.46]	[2.72]	[2.38]	[2.37]	[2.63]	[2.83]	[2.38]	[2.35]
	L/S	-0.0393	-0.0338	-0.0026	0.0006	-0.0499	-0.0376	-0.0054	-0.0024
		[-1.68]	[-2.09]	[-0.21]	[0.06]	[-2.05]	[-2.27]	[-0.45]	[-0.26]
SR	top	0.70	0.79	0.84	0.82	0.66	0.75	0.81	0.79
	bottom	0.82	0.88	0.76	0.75	0.87	0.90	0.78	0.75
	L/S	-0.51	-0.55	-0.06	0.01	-0.61	-0.59	-0.11	-0.06
		ESGC				strict ESGC			
mean	top	0.1391	0.1383	0.1434	0.1402	0.1341	0.1322	0.1392	0.1374
		[2.59]	[2.59]	[2.72]	[2.60]	[2.44]	[2.49]	[2.61]	[2.52]
	bottom	0.1751	0.1516	0.1356	0.1282	0.1717	0.1540	0.1388	0.1286
		[2.66]	[2.67]	[2.37]	[2.26]	[2.64]	[2.70]	[2.41]	[2.24]
	L/S	-0.0310	-0.0117	0.0070	0.0108	-0.0325	-0.0191	0.0004	0.0079
		[-1.38]	[-0.80]	[0.51]	[1.02]	[-1.49]	[-1.35]	[0.03]	[0.70]
SR	top	0.82	0.80	0.80	0.80	0.77	0.77	0.77	0.77
	bottom	0.91	0.84	0.75	0.72	0.89	0.85	0.77	0.72
	L/S	-0.41	-0.21	0.15	0.27	-0.43	-0.34	0.01	0.18
		ENV				strict ENV			
mean	top	0.1137	0.1289	0.1323	0.1364	0.1091	0.1269	0.1298	0.1335
		[2.20]	[2.40]	[2.59]	[2.64]	[2.05]	[2.35]	[2.49]	[2.54]
	bottom	0.1789	0.1677	0.1416	0.1344	0.1808	0.1666	0.1462	0.1352
		[2.69]	[2.77]	[2.32]	[2.27]	[2.63]	[2.79]	[3.23]	[2.29]
	L/S	-0.0561	-0.0337	-0.0082	0.0018	-0.0615	-0.0345	-0.0145	-0.0015
		[-2.14]	[-2.02]	[-0.59]	[0.18]	[-2.23]	[-2.14]	[-0.96]	[-0.14]
SR	top	0.69	0.76	0.81	0.82	0.64	0.74	0.79	0.80
	bottom	0.87	0.87	0.76	0.73	0.87	0.87	0.79	0.74
	L/S	-0.66	-0.54	-0.16	0.05	-0.72	-0.53	-0.27	-0.03
		SOC				strict SOC			
mean	top	0.1099	0.1275	0.1355	0.1375	0.1089	0.1255	0.1306	0.1357
		[2.36]	[2.71]	[2.64]	[2.62]	[2.35]	[2.62]	[2.50]	[2.56]
	bottom	0.1471	0.1338	0.1353	0.1372	0.1472	0.1360	0.1389	0.1406
		[2.31]	[2.35]	[2.44]	[2.42]	[2.30]	[2.41]	[2.47]	[2.48]
	L/S	-0.0328	-0.0056	0.0001	0.0003	-0.0338	-0.0093	-0.0073	-0.0043
		[-1.18]	[-0.3]	[0.01]	[0.04]	[-1.21]	[-0.54]	[-0.76]	[-0.56]
SR	top	0.75	0.82	0.82	0.81	0.74	0.80	0.78	0.79
	bottom	0.81	0.76	0.76	0.76	0.82	0.79	0.79	0.78
	L/S	-0.42	-0.09	0.00	0.01	-0.42	-0.16	-0.16	-0.12
		GOV				strict GOV			
mean	top	0.1325	0.1240	0.1306	0.1328	0.1224	0.1176	0.1296	0.1322
		[2.66]	[2.38]	[2.44]	[2.47]	[2.40]	[2.29]	[2.45]	[2.47]
	bottom	0.1463	0.1540	0.1386	0.1290	0.1539	0.1511	0.1386	0.1284
		[2.44]	[2.48]	[2.33]	[2.21]	[2.69]	[2.55]	[2.36]	[2.19]
	L/S	-0.0122	-0.0263	-0.0071	0.0034	-0.0276	-0.0295	-0.0081	0.0034
		[-0.62]	[-1.37]	[-0.58]	[0.39]	[-1.20]	[-1.63]	[-0.66]	[0.37]
SR	top	0.80	0.74	0.78	0.79	0.75	0.70	0.78	0.78
	bottom	0.78	0.80	0.75	0.70	0.83	0.81	0.76	0.71
	L/S	-0.16	-0.40	-0.15	0.09	-0.35	-0.48	-0.17	0.09

Table A.3: 4-factor model for the value-weighted long-short portfolio strategies

This table presents the 4-factor model for the value-weighted long-short strategies. The long-short strategy holds a long position in the top-rated ESG stocks and short position in the bottom-rated ESG stocks. The full sample strategies as well as strict alternatives with negative screening are reported for the portfolio cut-offs of 5%, 10%, 20% and 30%. All models are estimated over the sample period from January 2003 to May 2017 on a monthly basis. All portfolios are weighted by the firms' market capitalization. Adjusted R^2 is reported. The standard errors are estimated using the Newey and West (1987) method.

	ESG				strict ESG			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0035**	-0.0025*	-0.0004	0.0001	-0.0047**	-0.0025*	-0.0011	-0.0003
	(0.0017)	(0.0015)	(0.0011)	(0.001)	(0.002)	(0.0014)	(0.001)	(0.001)
Rm-rf	-0.0012	-0.0662	-0.0744*	-0.0689*	0.0461	-0.0063	-0.0389	-0.0408
	(0.0677)	(0.0670)	(0.0440)	(0.0404)	(0.0641)	(0.0608)	(0.0373)	(0.0388)
SMB	-0.317***	-0.317***	-0.355***	-0.314***	-0.359***	-0.280***	-0.350***	-0.299***
	(0.0858)	(0.0937)	(0.0592)	(0.0518)	(0.0838)	(0.0907)	(0.0604)	(0.0592)
HML	0.288***	0.245**	0.199***	0.167***	0.195***	0.181**	0.165***	0.134**
	(0.0704)	(0.0962)	(0.0604)	(0.0613)	(0.0673)	(0.0891)	(0.0596)	(0.0643)
WML	0.102*	0.0588	0.0432*	0.0362*	0.0925	-0.0255	0.00577	0.0083
	(0.0562)	(0.0544)	(0.0255)	(0.0212)	(0.0582)	(0.0393)	(0.0222)	(0.0135)
R^2	0.1124	0.1712	0.3026	0.2924	0.0884	0.1109	0.2564	0.2331
	ESGC				strict ESGC			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0053***	-0.0028**	-0.0015*	-0.0016	-0.0049**	-0.0032**	-0.0017*	-0.0017*
	(0.0019)	(0.0014)	(0.0008)	(0.001)	(0.0022)	(0.0014)	(0.001)	(0.001)
Rm-rf	-0.0439	0.0231	0.0117	-0.0406	-0.0241	0.0166	0.0127	-0.0344
	(0.0902)	(0.0393)	(0.0410)	(0.0265)	(0.0921)	(0.0427)	(0.0402)	(0.0273)
SMB	-0.135	-0.0702	-0.0679	-0.0679	-0.136	-0.0696	-0.0881	-0.0942*
	(0.0949)	(0.0750)	(0.0586)	(0.0479)	(0.111)	(0.0787)	(0.0564)	(0.0524)
HML	0.0894	0.165***	0.0838	0.205****	-0.0252	0.147**	0.0563	0.174***
	(0.0875)	(0.0577)	(0.0575)	(0.0472)	(0.0974)	(0.0583)	(0.0535)	(0.0578)
WML	0.0374	0.0480	0.0326*	0.0375*	0.0292	0.0373	0.0129	0.0327
	(0.0509)	(0.0314)	(0.0180)	(0.0216)	(0.0585)	(0.0306)	(0.0172)	(0.0226)
R^2	0.0066	0.0202	0.0042	0.1062	-0.0005	0.0062	-0.0032	0.0754
	ENV				strict ENV			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0028	-0.0024*	-0.0009	-0.006	-0.002	-0.0022	-0.0011	-0.006
	(0.0021)	(0.0015)	(0.0012)	(0.0011)	(0.0023)	(0.0018)	(0.0014)	(0.0011)
Rm-rf	-0.149	-0.0632	-0.0197	-0.0266	-0.175	-0.0868	0.00269	-0.0266
	(0.100)	(0.0453)	(0.0435)	(0.0376)	(0.107)	(0.0552)	(0.0457)	(0.0376)
SMB	-0.418***	-0.320***	-0.288***	-0.276***	-0.429***	-0.320***	-0.296***	-0.276***
	(0.106)	(0.110)	(0.0519)	(0.0543)	(0.113)	(0.120)	(0.0535)	(0.0543)
HML	0.0627	0.0553	0.0171	-0.00742	0.0495	-0.0007	-0.0130	-0.0074
	(0.120)	(0.0427)	(0.0516)	(0.0514)	(0.156)	(0.0646)	(0.0643)	(0.0514)
WML	0.0537	0.0532	0.0615**	0.0291	0.0290	0.0288	0.0297	0.0291
	(0.0481)	(0.0342)	(0.0264)	(0.0195)	(0.0547)	(0.0343)	(0.0241)	(0.0195)
R^2	0.2096	0.171	0.2000	0.2324	0.2138	0.1665	0.1577	0.2324
	SOC				strict SOC			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0015	0.0000	-0.0011	0.0000	-0.0014	-0.0006	-0.0018	-0.0008
	(0.0021)	(0.0015)	(0.0012)	(0.0012)	(0.0022)	(0.0014)	(0.0011)	(0.001)
Rm-rf	-0.106	-0.0976	-0.0688	-0.0725	-0.110	-0.0899	-0.0262	-0.0446
	(0.0835)	(0.0640)	(0.0512)	(0.0532)	(0.0920)	(0.0645)	(0.0417)	(0.0424)
SMB	-0.458***	-0.411***	-0.291***	-0.315***	-0.452***	-0.400***	-0.295***	-0.292***
	(0.134)	(0.0824)	(0.0627)	(0.0507)	(0.139)	(0.0845)	(0.0592)	(0.0511)
HML	0.238**	0.217	0.193**	0.155**	0.211*	0.202	0.162**	0.118*
	(0.102)	(0.146)	(0.0793)	(0.0754)	(0.112)	(0.148)	(0.0680)	(0.0639)
WML	0.0300	0.00941	0.0287	0.0285	-0.00480	-0.0237	0.00276	0.0279*
	(0.0413)	(0.0428)	(0.0362)	(0.0246)	(0.0375)	(0.0356)	(0.0186)	(0.0150)
R^2	0.1766	0.2517	0.2263	0.2905	0.1545	0.2313	0.2033	0.2642
	GOV				strict GOV			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	0.0015	-0.0012	0.0003	0.0012	0.005	-0.003*	-0.0002	0.001
	(0.0023)	(0.0016)	(0.0012)	(0.001)	(0.0018)	(0.0017)	(0.0011)	(0.0009)
Rm-rf	-0.156**	-0.0461	-0.139***	-0.135***	-0.0894*	0.00207	-0.109***	-0.112***
	(0.0697)	(0.0390)	(0.0440)	(0.0290)	(0.0514)	(0.0369)	(0.0394)	(0.0271)
SMB	-0.297**	-0.295***	-0.251***	-0.203***	-0.380***	-0.322***	-0.237***	-0.192***
	(0.119)	(0.0936)	(0.0631)	(0.0494)	(0.130)	(0.0942)	(0.0667)	(0.0570)
HML	0.426***	0.302***	0.273***	0.213***	0.310**	0.234***	0.224***	0.168***
	(0.123)	(0.0894)	(0.0737)	(0.0512)	(0.122)	(0.0845)	(0.0753)	(0.0531)
WML	0.0654	0.0893**	0.0649**	0.0402	0.0331	0.0343	0.0123	-0.0006
	(0.0709)	(0.0439)	(0.0319)	(0.0253)	(0.0724)	(0.0471)	(0.0261)	(0.0200)
R^2	0.2139	0.1666	0.2782	0.2902	0.1695	0.1264	0.2194	0.2325

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A.4: 4-factor model for the equal-weighted long-short portfolio strategies

This table presents the 4-factor model for the equal-weighted long-short strategies. The long-short strategy holds a long position in the top-rated ESG stocks and short position in the bottom-rated ESG stocks. The full sample strategies as well as strict alternatives with negative screening are reported for the portfolio cut-offs of 5%, 10%, 20% and 30%. All models are estimated over the sample period from January 2003 to May 2017 on a monthly basis. All portfolios are equally weighted. Adjusted R^2 is reported. The standard errors are estimated using the Newey and West (1987) method.

	ESG				strict ESG			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0023	-0.0016*	0.0008	0.0008	-0.0033**	-0.0022**	0.0004	0.0005
	(0.0015)	(0.0008)	(0.0006)	(0.0005)	(0.0016)	(0.001)	(0.0007)	(0.0005)
Rm-rf	-0.0292	-0.0606*	-0.0293	-0.0251	-0.0076	-0.0281	-0.0130	-0.0249
	(0.0333)	(0.0332)	(0.0230)	(0.0205)	(0.0382)	(0.0335)	(0.0262)	(0.0239)
SMB	-0.349***	-0.377***	-0.355***	-0.270***	-0.402***	-0.398***	-0.359***	-0.268***
	(0.0947)	(0.0670)	(0.0391)	(0.0387)	(0.101)	(0.0670)	(0.0440)	(0.0419)
HML	0.0473	0.0492	0.0414	0.0350	0.0307	0.0371	0.0347	0.0302
	(0.0588)	(0.0472)	(0.0326)	(0.0323)	(0.0597)	(0.0513)	(0.0330)	(0.0353)
WML	0.130***	0.0602**	0.0369*	0.0077	0.121***	-0.00401	-0.0028	-0.0135
	(0.0362)	(0.0263)	(0.0201)	(0.0234)	(0.0386)	(0.0414)	(0.0326)	(0.0312)
R^2	0.2068	0.3165	0.4026	0.3039	0.1966	0.2455	0.3447	0.2759
	ESGC				strict ESGC			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0018	-0.0005	0.0009	0.0013	-0.002	-0.0011	0.0003	0.001
	(0.0019)	(0.0012)	(0.001)	(0.0008)	(0.0019)	(0.0012)	(0.0011)	(0.0009)
Rm-rf	-0.0383	-0.0152	-0.0114	-0.0244	-0.0281	-0.0121	-0.0060	-0.0138
	(0.0612)	(0.0353)	(0.0224)	(0.0202)	(0.0614)	(0.0391)	(0.0244)	(0.0214)
SMB	-0.203**	-0.169***	-0.115**	-0.102**	-0.212**	-0.185***	-0.124**	-0.112**
	(0.0912)	(0.0596)	(0.0500)	(0.0401)	(0.0996)	(0.0658)	(0.0585)	(0.0455)
HML	0.0020	0.0149	0.0138	0.0198	-0.0337	0.0130	-0.0007	0.0142
	(0.0553)	(0.0495)	(0.0450)	(0.0458)	(0.0605)	(0.0536)	(0.0544)	(0.0532)
WML	0.0893**	-0.003	-0.0231	-0.0291	0.0640*	0.0097	-0.0511	-0.0552
	(0.0369)	(0.0242)	(0.0539)	(0.0417)	(0.0354)	(0.0254)	(0.0701)	(0.0551)
R^2	0.0877	0.0435	0.0258	0.0424	0.0723	0.0547	0.0353	0.0558
	ENV				strict ENV			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0035*	-0.0021*	0.0003	0.001	-0.0038*	-0.0021*	-0.0003	0.0006
	(0.0019)	(0.0011)	(0.0007)	(0.0006)	(0.002)	(0.0011)	(0.0009)	(0.0007)
Rm-rf	-0.0498	0.0185	-0.0398	-0.0291	-0.0640	-0.000945	-0.0267	-0.0167
	(0.0562)	(0.0376)	(0.0250)	(0.0205)	(0.0579)	(0.0386)	(0.0286)	(0.0211)
SMB	-0.398***	-0.380***	-0.307***	-0.263***	-0.418***	-0.366***	-0.314***	-0.258***
	(0.0711)	(0.0756)	(0.0556)	(0.0474)	(0.0770)	(0.0886)	(0.0646)	(0.0474)
HML	-0.123	-0.110***	-0.0576*	-0.0597**	-0.0846	-0.117***	-0.0690*	-0.0588**
	(0.0749)	(0.0396)	(0.0305)	(0.0275)	(0.0852)	(0.0396)	(0.0388)	(0.0266)
WML	0.111**	0.0749***	0.0231	0.0100	0.0781*	0.0203	0.0036	-0.0076
	(0.0431)	(0.0249)	(0.0422)	(0.0192)	(0.0448)	(0.0299)	(0.0463)	(0.0303)
R^2	0.2701	0.3025	0.3068	0.3451	0.2402	0.2370	0.2614	0.2743
	SOC				strict SOC			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	-0.0015	0.0003	0.0006	0.0006	-0.0015	0.0000	0.0000	0.0000
	(0.0018)	(0.0011)	(0.0007)	(0.0005)	(0.0019)	(0.0012)	(0.0007)	(0.0005)
Rm-rf	-0.0336	0.00570	-0.0116	-0.00712	-0.0464	0.0129	-0.00196	-0.00356
	(0.0452)	(0.0337)	(0.0257)	(0.0223)	(0.0500)	(0.0371)	(0.0272)	(0.0239)
SMB	-0.458***	-0.400***	-0.262***	-0.207***	-0.439***	-0.400***	-0.263***	-0.196***
	(0.0855)	(0.0614)	(0.0522)	(0.0430)	(0.0868)	(0.0664)	(0.0531)	(0.0449)
HML	-0.0585	0.0185	0.0197	-0.0127	-0.0463	0.0223	0.0113	-0.0130
	(0.0624)	(0.0556)	(0.0359)	(0.0248)	(0.0616)	(0.0583)	(0.0332)	(0.0227)
WML	0.0820***	0.0492**	0.0024	0.0082	0.0504	0.0200	-0.0183	-0.0035
	(0.0234)	(0.0210)	(0.0232)	(0.0192)	(0.0311)	(0.0246)	(0.0316)	(0.0297)
R^2	0.2853	0.2823	0.2055	20.95	0.2409	0.2493	0.1977	0.1695
	GOV				strict GOV			
	ls5	ls10	ls20	ls30	ls5	ls10	ls20	ls30
Alpha	0.00001	-0.0014	0.0004	0.0011*	-0.0013	-0.0018	0.0003	0.0011*
	(0.0013)	(0.0013)	(0.0009)	(0.0006)	(0.002)	(0.0014)	(0.0009)	(0.0006)
Rm-rf	-0.0884*	-0.0391	-0.0736**	-0.0652***	-0.0619	-0.0146	-0.0639**	-0.0558***
	(0.0478)	(0.0380)	(0.0292)	(0.0188)	(0.0686)	(0.0392)	(0.0255)	(0.0188)
SMB	-0.171*	-0.239***	-0.194***	-0.177***	-0.240***	-0.276***	-0.210***	-0.185***
	(0.0885)	(0.0810)	(0.0475)	(0.0340)	(0.0857)	(0.0676)	(0.0468)	(0.0382)
HML	0.118	0.0746	0.0659	0.102**	0.0404	0.00562	0.0363	0.0824***
	(0.0995)	(0.109)	(0.0812)	(0.0427)	(0.0875)	(0.0850)	(0.0602)	(0.0278)
WML	0.0637	0.135***	0.0325	0.0279*	0.0523	0.0709**	-0.0093	0.0017
	(0.0757)	(0.0348)	(0.0198)	(0.0168)	(0.0712)	(0.0285)	(0.0230)	(0.0228)
R^2	0.0800	0.2005	0.1921	0.2633	0.0837	0.1602	0.1746	0.2213

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Appendix B

ESG database score measures

Table B.1: Category scores

The table lists the category scores, the pillar they are connected with and their description. This information is obtained from Thomson returns ESG scores report (Thomson Reuters, 2017).

Score	Pillar	Definition
Resource Use Score	ENV	The Resource Use Score reflects a company's performance and capacity to reduce the use of materials, energy or water, and to find more eco-efficient solutions by improving supply chain management.
Emissions Score	ENV	The Emission Reduction Score measures a company's commitment and effectiveness towards reducing environmental emission in the production and operational processes.
Innovation Score	ENV	The Innovation Score reflects a company's capacity to reduce the environmental costs and burdens for its customers, and thereby creating new market opportunities through new environmental technologies and processes or eco-designed products.
Workforce Score	SOC	The Workforce Score measures a company's effectiveness towards job satisfaction, a healthy and safe workplace, maintaining diversity and equal opportunities, and development opportunities for its workforce.
Human Rights Score	SOC	The Human rights category score measures a company's effectiveness towards respecting the fundamental human rights conventions.
Community Score	SOC	The Community Score measures the company's commitment towards being a good citizen, protecting public health and respecting business ethics.
Product Responsibility Score	SOC	The Product Responsibility Score reflects a company's capacity to produce quality goods and services integrating the customer's health and safety, integrity and data privacy.
Management Score	GOV	The Management Score measures a company's commitment and effectiveness towards following best practice corporate governance principles.
Shareholders Score	GOV	The Shareholders Score measures a company's effectiveness towards equal treatment of shareholders and the use of anti-takeover devices.
CSR Strategy Score	GOV	The CSR Strategy Score reflects a company's practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes.

Table B.2: Controversy measures

This table lists controversy measures that are collected for the computation of the ESG Controversy Category Score. This information is obtained from Thomson returns ESG scores report (Thomson Reuters, 2017).

Category	Controversy
Community	Anti-Competition Controversy
Community	Business Ethics Controversies
Community	Intellectual Property Controversies
Community	Critical Countries Controversies
Community	Public Health Controversies
Community	Tax Fraud Controversies
Human Rights	Child Labor Controversies
Human Rights	Human Rights Controversies
Management	Mgt Compensation Controversies Count
Product Responsibility	Consumer Controversies
Product Responsibility	Controversies Customer Health & Safety
Product Responsibility	Controversies Privacy
Product Responsibility	Controversies Product Access
Product Responsibility	Controversies Responsible Marketing
Product Responsibility	Controversies Responsible R&D
Resource Use	Environmental Controversies
Shareholders	Accounting Controversies Count
Shareholders	Insider Dealings Controversies Count
Shareholders	Shareholder Rights Controversies Count
Workforce	Diversity and Opportunity Controversies
Workforce	Employees Health & Safety Controversies
Workforce	Wages Working Condition Controversies Count
Workforce	Management Departures