On Responsible Investment: Generating Abnormal Returns with Screening Strategies

Luuk te Grotenhuis[a]

Thesis supervisor: Gabriele Lepori[b]

[a]) MSc. Finance & Strategic Management, Copenhagen Business School (Denmark)

[b]) Institut for Finansiering, Copenhagen Business School

August 2012

No. of characters: 125,658
(55.23 standard pages, excluding tables and figures)

Abstract

This thesis provides evidence contradicting recent studies that claim excess risk-adjusted returns can be generated by forming portfolios on extra-financial information. Three screening strategies based on environmental, social and governance (ESG) indicators are empirically tested for their ability to achieve abnormal returns over the 2004-2011 period.

Responsible investment can be subdivided according to either values-driven or profit-seeking investors. They have diverging motives and are respectively served by negative or positive screening strategies. It is conjectured that employing negative, values-driven screening will result in underperformance. On the other hand, positive, profit-seeking screening strategies should display outperformance. Based on MSCI ESG STATS rating data, high- and low-rated stock portfolios are formed and consequently tested for abnormal returns with the CAPM, Fama-French three-factor, and Carhart four-factor models.

The results for a negative screening strategy that excludes stocks of companies in perceived controversial business areas hint towards underperformance, with a near to significant alpha value of -2.64 percent (p-value: 0.11). In contrast to earlier research, positive screening strategies only incorporating the best-performing companies on ESG indicators also exhibit underperformance. The statistically strongest result is found for the best-in-class strategy, which maintains a balanced sector allocation in the portfolio. The alpha value of -2.85 percent (p-value 0.02) falsifies the assumption that excess risk-adjusted returns can be generated by arranging portfolios on ESG information. As a result, this thesis did not find a positive link between corporate social performance and corporate financial performance.

Now responsible investment is on its way to become a mainstream method of investing, the results in this thesis question whether profit-seeking investors should invest their money according to extra-financial information. Evidenced by the underperformance of positive screening strategies, it seems that money cannot be put to both socially and financially 'good' use.

Table of content

Table of content (continued)

1. Introduction

1.1 Problem Indication

Over the past decades, responsible investment has been growing enormously. Its special feature, combining financial and social rationales, and the growing awareness amongst investors for social, environmental and governance (ESG) issues, mainly spurs this trend. However, integrating non-financial factors into investment decision-making contradicts the traditional finance theories. When the mean-variance criterion dictates the optimal portfolio allocation, basing investment decisions on any other indicators would go at the expense of expected returns. In this view, incorporating social factors would unnecessarily limit the investment universe and should not offer a competitive edge in establishing excess risk-adjusted returns (Bodie, Kane & Marcus, 2009; Renneboog, Ter Horst & Zhang, 2008)

A whole stream of research has studied the performance effects of socially screened portfolios over the years. However, most results are based on mutual funds data and often were inconclusive; no consensus has been reached. In recent years, it has been hypothesised that the mixed results might be attributable to different investor motivations. Some investors predominantly want to screen their portfolio to refrain from certain stocks, while others wish to use the non-financial indicators to build portfolios that they believe will outperform the market. Both logics apply their own strategy to reach that goal; the former is best served by negative, exclusionary screening, while the latter is likely to use positive screening. Not controlling for these different motivations and screening strategies might very well be the reason behind the inconclusive evidence in most studies on the performance of responsible investment (Bauer, Koedijk & Otten, 2005; Derwall, Koedijk & Ter Horst, 2011).

Over the past five years, instead of using mutual funds data, academic research has used self-constructed stock portfolios based on sustainability ratings. By using this approach, the scholar chooses which stocks to in- and exclude in the analysis. This better enables researchers to control and study the different sources of abnormal returns. It is theorised that the investors who use negative screens are driven by ethical values and not solely financial benefits, which according to the '*Doing Good, but Not Well*' hypothesis can lead to long-term, sustained *under*performance. The other group of profit-seeking investors, applying positive screening methods, acts according to a '*Doing Good while Doing Well*' hypothesis. For this hypothesis, ESG factors are considered as extra-financial information and the basis for *positive* excess risk-adjusted returns. This assumes a link between corporate social performance and financial performance. Moreover, the impact of the strong social performance has to be unexpected for it to yield abnormal returns. Therefore, it is can be conjectured that a long-term sustained excess risk-adjusted return cannot be created by

incorporating ESG criteria, but that the effect is temporary in nature. Nevertheless, foregoing research found some ESG indicators and strategies that delivered significant positive abnormal returns over a prolonged period (Derwall, et al., 2011; Kempf & Osthoff, 2007; Statman & Glushkov, 2009).

1.2 Problem Statement

Considering all this existing evidence, it begs the question whether social screening strategies can, or cannot achieve abnormal positive returns for responsible investors. In addition, the different motivations and screening methods suggest that conclusions on the financial performance should not be generalised for a single, homogeneous group of responsible investors. Rather, any definitive statement on the performance of responsible investment must be segmented according to the underlying rationales and drivers, either being a values-driven or a profit-seeking motive.

This thesis sets out to study the effects of negative and positive screening strategies and the possibility to achieve abnormal returns by incorporating ESG indicators into the portfolio selection. The results will contribute to existing research by explicitly taking into account the segmentation in responsible investment motivations. Furthermore, the stock universe under analysis is extended by including the 3,000 largest listed companies in the United States. Lastly, the time span ranging from 2004 to 2011 covers the most recent years, which are often not included in the other studies.

1.3 Research Question

The impact of incorporating non-financial performance indicators into investment decision-making and portfolio allocation is studied. The conjecture is that investors use portfolio screens for different reasons and therefore those different screening strategies will have a mixed impact on financial performance and the ability to achieve abnormal returns. Summarised, the research question for this thesis is:

What are the effects of negative and positive ESG-based screening strategies on achieving excess risk-adjusted portfolio returns?

This question is empirically tested by constructing stock portfolios according to several screening techniques. Consequently, the existence of abnormal returns – or alpha – for these portfolios is investigated for the years 2004-2011. For this purpose, the following hypotheses have been put forth:

Hypothesis 1: *A low-rated portfolio with stocks of companies involved in at least one controversial business area will outperform a high-rated portfolio consisting of all other stocks.*

Hypothesis 2: *A high-rated portfolio with stocks of companies with high ESG ratings will outperform a low-rated portfolio with stocks of companies with low ESG ratings.*

Hypothesis 3: *A high-rated portfolio with stocks of companies with high best-in-class ESG ratings will outperform a low-rated portfolio with stocks of companies with low best-in-class ESG ratings.*

1.4 Overview of the Thesis

The remainder of this thesis will start by sketching the historical development of responsible investment. Furthermore, chapter 2 also summarises some market statistics on responsible investment in the United States. The literature review, in chapter 3, covers the first part of the hypothetic-deductive analysis in this thesis. It defines what is included in the definition of responsible investment and reviews previous performance studies. Chapter 4 presents the methodology and data sources used in the empirical analysis. Moreover, the hypotheses are further introduced. Consequently, the results of the hypothesis testing are summarised in chapter 5, complemented with some additional subsample tests. Lastly, the conclusions, limitations, and recommendations can be found in chapter 6 (Sekaran, 2003).

2. Market Overview

2.1 Historical Development of Responsible Investment

From a historical perspective, the responsible use of capital has been promoted for many centuries. Various studies highlight the influence of religious movements like Christianity, Judaism, and the Islam on financial transactions. Their traditions are built on embracing peace and not harming fellow humans, values that they also integrate in business dealings. An often quoted example of how these beliefs affect investments is the prohibition of – disproportionate – interest on loans. The early origins in the United States, the most well established institutional market for responsible investment, are traced back to the seventeenth century. The Quaker and Methodism societies restricted their investments by banning any trades that aided the Civil War or were used for the misappropriation of others – i.e. slavery. Nowadays, these deep historic roots are still obvious in the United States. They are most prominently displayed by the use of exclusionary screens, avoiding investments in companies that produce alcohol, offer gambling services or whose business goal is not aligned with fundamental values and beliefs; the so-called 'sin stocks' (Blowfield & Murray, 2008; Renneboog et al., 2008; Schueth, 2003).

The evolution of modern-day socially responsible investing (SRI) started from the 1960s onwards. Growing public sentiments against the Vietnam War, South-African 'Apartheid', and nuclear power had a profound effect on investors' preferences. Individual investors were confronted with the fact that their investment choices could have substantial social consequences. Responsible investing no longer solely meant the inclusion of fundamental ethical and religious beliefs. In contrast, investors started to select investments according to their own idiosyncratic thoughts and opinions. Besides this changing attitude and broader interpretation of responsible investing, the increasing amount of retail investment funds available to private investors also supported the birth and progression of modern-day SRI. The Pax World Fund, established in 1971, is an example of such a fund that refrained from investing in companies that profited from the Vietnam War.

Since then several milestones were accomplished. In 1983, the Ethical Investment Research Service (EIRIS) was established, which started publishing research on companies concerning their social performance. In 1984, the first U.S. Social Investment Forum (U.S. SIF) survey was conducted; totalling the U.S. market for SRI at $40 billion and showing a growing trend. This increase was further spurred by a focus on environmental sustainability and accountability during the late 1980s, mainly caused by large mishaps like the Chernobyl nuclear accident and the Exxon Valdez oil spill. In the following years, responsible investment continuously grew to become

a mainstream method of investment. The introduction of the FTSE4Good index in 2001 and the availability of the first socially screened exchange-traded funds (ETFs) in 2005 underline this advancement. At present, it can be concluded that responsible investment has grown from a marginal, virtually charitable phenomenon to a full-fledged investment alternative for the institutionalised and mainstream investor (Blowfield & Murray, 2008; Jansson & Biel, 2011; Renneboog et al., 2008).

At the supranational level, Europe had set standards for corporate social responsibility and sustainability by means of the 2000 European Union Lisbon Agenda. Two years later, the World Summit on Sustainable Development in Johannesburg, designed the 'London Principles' specifically for financial institutions, stimulating them to advance their efforts with regards to economic prosperity, environmental protection, and social development. Thus, besides the growing acceptance of responsible investment by market participants, also the legislative actors took interest in developing and institutionalising SRI. However, it took until 2006 for a universal standard of responsible investing to emerge. The United Nations Environment Programme's Finance Initiative (UNEP-FI) created the Principles for Responsible Investment[1] (UNPRI), which together with the UN Global Compact[2] norms offers a voluntary frame of reference for responsible investing. Together, these should encourage investors and investment managers to incorporate social, non-financial factors into their analysis, making leeway for worldwide acceptance of responsible investing (Blowfield & Murray, 2008; UNPRI, 2012).

2.2 Market Statistics

The U.S. Social Investment Forum (US SIF, 2010) keeps record of all the socially responsible assets under management in the U.S. These data are published in a biennial report, the most recent one dating back to 2010; the 2012 version will be available as of November 2012.

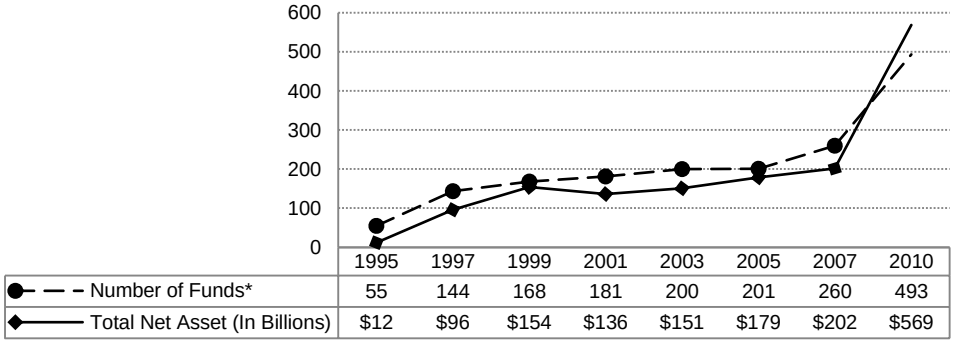
In 2010, out of all assets under professional management in the United States – in total $25.2 trillion as measured by ThomsonReuters – more than 12 percent was dedicated to responsible investment. This includes retail funds, but more decisively also institutional fund, like pension funds and separate accounts (Renneboog et al., 2008). The SRI assets grew from $639 billion in 1995 up to $3.07 trillion by 2010; an increase over this period of 380 percent. Thereby it outpaced the overall U.S. investment industry, which only grew 260 percent over the same period.

[1] See http://www.unpri.org/principles/

[2] See http://www.unglobalcompact.org/AboutTheGC/TheTenPrinciples/index.html

This growth in SRI assets sums up to a compound annual growth rate[3] of over 11% per year. Especially during the most recent crisis years, SRI has shown consistent double-digit growth, while "the overall universe of professionally managed assets has remained roughly flat" (US SIF, 2010; p.8). Figure 2a shows in further detail that the sharp incline since 2007 is also apparent in the investment funds. The number of funds that integrate principles for responsible investment in their management has risen steeply. The graph shows that investment vehicles increasingly take environmental, social, and governance (ESG) factors into consideration (US SIF, 2010).

Figure 2a. Investment funds including ESG-factors 1995-2010

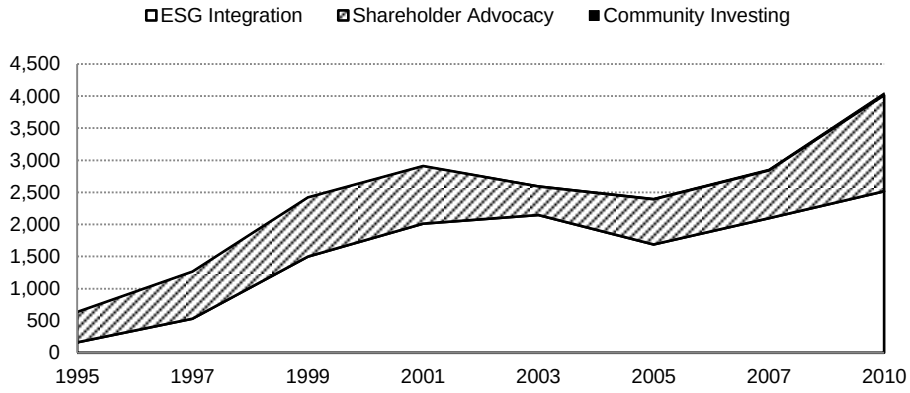


	1995	1997	1999	2001	2003	2005	2007	2010
Number of Funds*	55	144	168	181	200	201	260	493
Total Net Asset (In Billions)	$12	$96	$154	$136	$151	$179	$202	$569

**) Note:* Separate accounts are excluded from these data.
Source: U.S. SIF (2010) Report on Socially Responsible Investing Trends in the United States

Figure 2b shows the growth according to separate responsible investment strategies – which will be discussed in the upcoming chapter. The most significant increase is seen in the incorporation of environmental, social, and governance (ESG) issues into investment analysis and decision-making processes, outgrowing shareholder advocacy or active ownership (UNPRI, 2012).

Figure 2b. Socially responsible investing in the United States 1995-2010



Note: Totals in the graph are not corrected for overlapping strategies.
Source: U.S. SIF (2010) Report on Socially Responsible Investing Trends in the United States

[3] $Compound\ Annual\ Growth\ Rate\ (CAGR)$: $\left[\frac{\$639\ billion}{\$3.07\ trillion}^{1/15}\right] - 1 = 0{,}1103$

3. Literature Review

3.1 Corporate Social Responsibility and Corporate Financial Performance

Research on responsible investment is part of the broader academic study into corporate social responsibility (CSR). CSR has, analogous to responsible investment, seen a tremendous development since the 1960s (Epstein, 2004). It can be defined as "the notion that corporations have an obligation to constituent groups in society other than stockholders and beyond that prescribed by law and union contract" (Jones, 1980; as cited in Carroll, 1999; p.284). The concept of CSR does not only cover social community issues, but it should be broadly defined, including social, ethical, and environmental concerns. This 'stakeholder' approach is a voluntary philosophy "beyond the narrow economic, technical, and legal requirements of the corporation and thus begins where the law ends" (Vallentin, 2003; p. 259).

The stakeholder approach has been a subject of heavy academic debate with a lot of research looking into the relationship between corporate social performance (CSP) and corporate financial performance (CFP). Friedman (1962) claims that every dollar invested in CSR initiatives is a wasted dollar. His classical argument against CSR states that companies and the economic environment are best served by profit maximization. Any business manager that deviates from this goal neglects its fiduciary duty towards the owner of the company, the shareholder. Jensen (2001) adds to this argument that it would be too complex to incorporate all stakeholders and their particular issues into decision-making.

Freeman (1984) does see the benefits of CSR. He claims that profit maximization does not deliver the desired outcome in the real world, where conflicts of interest, information asymmetry, and costs influence transactions and prohibit an optimal distribution. In his view, CSR can be an ultimate strategy to align the interests of various stakeholders and a way to minimize these transaction costs.

In a meta-analysis Orlitzky, Schmidt and Rynes (2003) have compiled 52 previous quantitative studies investigating the CSP-CFP link. Up until that point, most evidence was stated to be too scattered to make generalizable inferences. By rigorously reviewing all results, they were able to draw a definitive conclusion across studies. Orlitzky et al. found that CSP is positively – bidirectional and simultaneously – correlated with CFP. The mixed interpretation of the evidence in the preceding years was attributed to sampling and measurement errors. The positive correlation was strongest for accounting measures and less strong for market-based metrics, which were not heavily investigated at that time.

Becchetti and Ciciretti (2009) specifically studied market-based CFP of socially responsible companies and concluded that a stakeholder-driven approach is not a 'free lunch' for corporations. They saw that CSR initiatives caused a significant redistribution of wealth from stockholders to the welfare of a broader set of corporate stakeholders. According to Becchetti and Ciciretti this reallocation of value unavoidably leads to lower stock returns (CFP).

The CSP-CFP relationship will not be revisited in the remainder of this thesis, but the existence of such a relationship will be implicitly assumed. By constructing portfolios based on top- and bottom performers on environmental, social and governance criteria, this thesis hopes to find a source of unexplained financial return.

3.2 Working towards a Definition of Responsible Investing

In the foregoing sections, the terms *socially responsible investing* (SRI), *ethical investing* (EI) and *responsible investing* (RI) have been used interchangeably. This is an often seen practice in the existing literature on responsible investing. Given its prolonged development, scholars have used different names over time. In an attempt to summarise the wide array of names used, Schueth (2003) lists the following titles: *social investing*, *socially aware/conscious investing*, *values-/mission-based investing* and *green investing*. These all describe a similar phenomenon and scholars do not explicitly make any distinction in investment style or meaning (Bauer et al., 2005).

For the purpose of this thesis, a clear distinction between the different names will be made. In the upcoming sections of this thesis, it will become clear that there is a significant difference between the pure ethical style and the more mainstream approach to responsible investing. Therefore, existing definitions will be reviewed and consequently segmented according to practical investment implementation.

In line with the description of modern-day SRI, as stated in the historical overview, Schueth (2003) delivers one of the broadest definitions: "the process of integrating personal values and societal concerns into investment decision-making" (p.190). The U.S. Social Investment Forum (US SIF; as cited in Geczy, Stambaugh & Levin, 2003; p.2) sharpens this broad characterization by putting emphasis on the conventional financial analysis. In U.S. SIF's opinion, a framework of solid financial analysis should be supplemented with the impact of investments on their social and environmental context, taking into account both positive and negative effects.

Vallentin (2003) also describes SRI as investments that combine the financial objectives of investors with their commitments towards social issues. In this respect, he defines 'social' as

matters regarding “peace, social justice, economic development, or a healthy environment” (p.257). He makes an important addition by stating that these investments do not necessarily need to deliver a financial return, but that the first priority is their impact on the society at large.

In a recent historic review of SRI, Blowfield and Murray (2008) work towards a practical interpretation of this often mentioned societal context. In their opinion, valuing the impact on the society should be done by integrating environmental, social and governance (ESG) concerns into the investment decision. However, the ultimate goal should always be to maximize the financial returns (the ‘bottom line’). They proclaim this integrated approach could lead to a new performance paradigm of risk, return, and *meaning*.

The concepts according to which a subdivision within the definition of responsible investment can be made are already apparent in the above-mentioned characterisations. Common denominators in the descriptions are: (1) the societal impact, (2) personal values, and (3) financial analysis. These constructs can be segmented according to investment approach. In order to make this segmentation, first the different strategies of responsible investment will be discussed.

Responsible Investment Strategies

In practice there is no standard method of responsible investing, moreover a responsible investor can choose which strategy to apply. The choice of strategy and subsequent approach enables an investor to accentuate one or more of these three common denominators[4], thereby creating his own distinctive investment scheme.

Overall, three separate strategies for responsible investment can be distinguished: *screening*, *shareholder advocacy*, and *community investment* (Blowfield & Murray, 2008). This thesis will focus on screening, since it can be applied in the pre-investment stage. At that time, the investment screens can be used to set up a trading strategy and form a corresponding portfolio. The other two responsible investment strategies rely on active engagement by the investor, respectively being an active owner and advocating for positive change, or participating in small focused projects, aiding less developed communities (Schueth, 2003).

When using investment screening, the choice of screening method separates ethical motives from financially driven goals. By either making use of negative or positive screening, the investor highlights which extra-financial information should be incorporated in the portfolio selection (Derwall et al., 2011). Subsequently, these different types of screening will be reviewed.

[4]
(1) Societal impact,
(2) personal values, and
(3) financial analysis.

Negative Screening

The negative approach means using exclusionary screens and the investors with religious agendas often apply it; shying away from the earlier mentioned 'sin stocks' of companies that operate in perceived controversial business areas. Tobacco (88%) is the most prominent negative screen, followed by alcohol (75%) and gambling (23%). These screens can be motivated in ethical beliefs and used as means of communicating to the general public. In the latter sense, excluding a company is an ethical statement that can help reaffirm the reputation of an investor (e.g. churches). 'Norms-based' screening, another form of negative screening, is not directly connected to true ethical investment motives. It excludes companies to comply with internationally agreed standards, like the UN Global Compact principles or the International Labour Organization standards.

By shunning companies from a portfolio, the specific risk can be reduced. If certain business areas are perceived as controversial, these might carry inherent risks. An often-mentioned example is litigation risk for tobacco companies, which might depress prices and valuations once a law suit has been filed. On the other hand, an obvious consequence of excluding some companies from the investment universe is a smaller investment set. With a high screening intensity this might alter the geographical and sector allocation and reduce risk-sharing benefits of portfolio diversification. The influence of screening practices will be examined in more detail in the upcoming chapters (Blowfield & Murray, 2008; Salaber, 2007; US SIF, 2012; Vallentin, 2003).

Positive Screening

A positive screening strategy does not lead to the exclusion of controversial companies. Rather, it strives to incorporate qualitative, non-financial information into investment decision-making, searching for the top performers on environmental, social and governance (ESG) criteria. The proponents of such affirmative screening methods often claim the existence of a commercial agenda or business case for responsible investment, arguing that the market does not correctly value corporate social responsibility (CSR) initiatives[5]. As a result, they state that specifically investing in these possibly undervalued companies will deliver positive financial returns, a 'triple bottom line' perspective.

'Best-in-class' screening is a distinct positive screening strategy. It also rates companies on a set of criteria, but always maintains a balanced sector allocation in the portfolio. The best performing companies within an industry are selected for investment, and no sectors are excluded upfront. This strategy overcomes exclusion of companies in relatively 'dirty' industries on

[5] Environmental, social, and governance (ESG) criteria and corporate social responsibility (CSR) are two sides of the same coin. Similar because they promote sustainable business practices, but from different perspectives since the former describes the investor-driven integration of social issues in investment decision-making and the latter is a business-driven initiative fulfilling social goals.

beforehand. Moreover, the businesses are judged on their performance in comparison with their peers. Therefore, the best-in-class strategy is often applied in indices, like the Dow Jones Sustainability Index (DJSI).

In summary, screening choices are always grey. Companies never are true top- or bottom performers on all examined ESG criteria. Hence, the perfect company does not exist (Vallentin, 2003; Kempf & Osthoff, 2007; Blowfield & Murray, 2008; SAM, 2012).

Responsible Investment Motivations

In addition to different strategies, investors are also found to have different investment motives, another parameter for segmentation. First, consider the difference between a responsible and a conventional investor. Both will invest in companies with positive net present value (NPV) to shareholders and a positive corporate social responsibility (CSR) record. Likewise, neither will invest in companies with negative NPV and negative CSR. In table 3, respectively A and D illustrate these cases. The difference arises with cases B and C. Conventional investors will primarily look at risk and return and therefore choose to invest in companies with a positive NPV, despite any negative CSR practices. On the other hand, the responsible investor is motivated by social objectives and could be willing to invest in a company with positive CSR, but showing financial underperformance (Renneboog et al., 2008).

Table 3. Segmentation of investment possibilities.

Companies	Positive NPV	Negative NPV
Positive CSR	(A) Both responsible *and conventional* investors invest	(B) only responsible investors (with positive screens) invest
Negative CSR	(B) only conventional investors invest	(D) neither *conventional* nor responsible investors invest

Adapted from: Renneboog, Ter Horst & Zhang (2008; p.1734)

Schueth (2003) describes the underlying motivations of responsible investors as (1) being able to feel good about the way their money is invested, aligning it with personal principles and concerns, and (2) by playing a role as agent of change, encouraging and establishing positive developments. Yet, Renneboog et al. (2008) draw a harder distinction. Besides the typical norms-constrained and/or aspirational motives, they believe a subgroup of responsible investors purely acts on the economic rational of wealth-maximization. Where the first group is driven by the act of social responsibility, from which they derive non-financial utility, the second group does not want to get involved with any form of charity or philanthropy and believes that positive returns due to CSR can be made. Therefore, responsible investors are not a homogeneous group. Rather, different motives exist which creates a heterogeneous group of investors, consisting of both values-driven investors and profit-seeking investors (Derwall et al.,2011).

'Shunned stock' and 'Errors-in-expectations' hypotheses

As illustrated in table 3, responsible investors are often associated with underperformance. If markets correctly value stock prices, responsible investments should generate weak returns. Firstly, because of underinvestment in positive NPV companies (i.e. negative screening) and secondly by overinvesting in negative NPV companies with positive ESG-scores (i.e. positive screening). In efficient markets public information on ESG issues should not offer a competitive edge to achieve abnormal returns. Nevertheless, the strong form of the efficient market hypothesis (EMH) is generally believed to be too extreme. Typically, even in highly competitive settings as financial markets only near-efficiency can be observed. This gives innovative and skilful investors opportunities to "generate better risk-adjusted returns" (Bodie et al., 2009; Renneboog et al., 2008; p.1734). In that sense, a thorough screening process could unlock value-relevant information unknown to the market. If the market has mispriced these stocks in the short-term, any benefits from CSR initiatives that materialize in the end can result in outperformance.

Based on this finding, Derwall et al. (2011) hypothesized that two effects might be simultaneously in play. Firstly, investors that form their holdings in line with personal values are more induced to shield controversial stocks from their portfolio; as a result underinvesting in positive NPV companies and therefore general underperformance. They call this the '*shunned stock*' hypothesis. An effect that should be mainly observed amongst ethically motivated, values-driven investors. The second effect focuses on investors that set out to find companies that derive a competitive advantage on ESG factors, not yet recognized by the market. This group of investors is called profit-seeking investors. A labour-intensive positive screening process might result in value-relevant information not priced by the market, consequently offering possibilities for outperformance, an effect which Derwall et al. named the '*errors-in-expectations*' hypothesis.

The next sections will review the evidence on the performance of responsible investment. Most of these studies did not yet incorporate the above outlined differences in investor motives and strategies, thereby delivering mixed results. For this thesis, the segmentation in values-driven and profit-seeking (value-driven) responsible investors will be used in the empirical analysis, for which negative screening policies serve the values-driven investor and profit-driven investors use positive screening methods.

3.3 Mutual Fund Studies on Performance of Responsible Investment

The growing popularity of responsible investment instantly raised ethical and legal questions, but surprisingly, empirical research into the performance effects lagged behind. In the

1980s, when academic interest caught on, its focus has been mainly on studying the returns[6] of the abovementioned screening strategies. Rudd (1981) was one of the first scholars to state that these strategies of investment targeting and exclusion cause a portfolio to be inherently biased. Building on arguments from modern portfolio theory, he notes that the extra risk created in a constrained, and thereby less diversified, portfolio is not compensated by a higher expected return. These insights were later popularly labelled the 'Markowitz view', denoting that investments incorporating social motives will become "by nature subsets of the market portfolio" (Bauer et al., 2005; p.1752).

Hamilton, Jo and Statman (1993) formally stated three hypotheses to further the studies into the performance of socially responsible funds. They hypothesized that relative to their conventional counterparts, the risk-adjusted returns of responsible funds are: (1) identical, (2) underperforming, or (3) outperforming. Statman and Glushkov (2009) retitled these as respectively 'No Effect', 'Doing Good but Not Well', and 'Doing Good while Doing Well'. In the first case, the market does not price the non-financial factors used to create the responsible portfolio and no differences in return should be observed. Alternatively, in the latter two hypotheses the market does price ESG-integration, either with negative or positive consequences.

Over the past twenty years, these assertions have been widely tested. Yet, early studies did not find any significant differences between the returns on responsible versus conventional portfolios (Bello, 2005; Guerard, 1997; Hamilton et al., 1993). Later, Statman (2000) concluded that socially screened mutual funds underperform compared to indices like the S&P500 and its socially responsible equivalent, the DSI400 (currently renamed to the 'MSCI KLD400'). However, he found the same to be true for conventional mutual funds. More importantly, at the fund level, like the other studies, no significant difference between conventional and responsible investment was found.

Studies on responsible investment in the U.K. have indicated a small cap bias in socially screened funds. In contrast to the before mentioned studies, a slight outperformance by responsible funds versus a market-wide index was initially reported. This evidence was supplemented with proof on a bias towards stocks of smaller size. Follow-up studies used different benchmarks and tests to advance these results. Ultimately, a modified two-factor Jensen's alpha model – incorporating a SMB-factor to account for company size – was used to confirm the small cap bias. Furthermore, with this revised market model it was concluded that, in line with Statman (2000), no significant difference between responsible and normal fund returns existed (Bauer et al., 2005).

[6] Investor returns; share price or share price appreciation, as a market-based measure of corporate financial performance (CFP).

Furthermore, according to a comprehensive review by Bauer et al. (2005) U.S. funds are exposed to large caps and growth stocks, which resulted in the conclusion that the outperformance of the socially responsible DSI index over the S&P500 was also due to style effects. Compounding the U.K and U.S. results, Bauer et al. claim that the non-financial social factors might have no effect at all and that performance is merely based on biases towards size- and sector allocations.

In general, the research that uses mutual funds exhibits a major pitfall. Since the studies are not focused on the individual mutual fund performance, often a group of mutual funds is combined in a portfolio. By forming such a portfolio, it is not possible to differentiate between the influences of the screening methods applied. Bauer et al. (2005) state that the mixed, inconclusive evidence of mutual fund studies can be attributed to different screening strategies used amongst fund managers, and that these varied strategies might cancel each other out (Derwall et al., 2011).

Norms-based screening is another factor that prohibits a good comparison between conventional and responsible funds. Since norms-based screening incorporates globally accepted norms and standards, some exclusion parameters that studies have used to select the sample of responsible funds can also be present among their sample of conventional funds. For example, the exclusion list of the Norwegian Governmental Pension Fund (GPF) has been widely used by conventional investment funds in Scandinavia (Bengtsson, 2008). Therefore, non-responsible funds are just as likely to exclude producers of, for instance, controversial weapons. Hence, it is not odd that a specific performance comparison on the effects of screening can deliver insignificant results, since both funds – responsible and conventional – might very well exclude similar assets. This conclusion is endorsed by Bello (2005), who concluded a broad review of existing literature by stating that responsible funds are not very different from conventional ones, when looking at portfolio holdings and the degree of diversification. After which he reaches the general conclusion that mutual fund investment performance is not significantly influenced by the use of social screens.

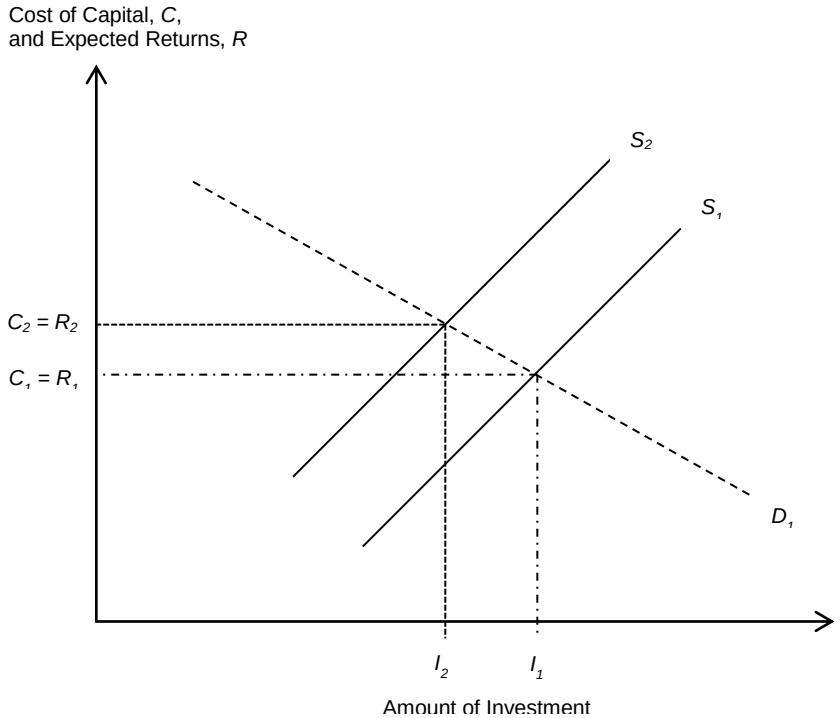
Lastly, by comparing performance over different mutual funds the scholars were not exclusively comparing responsible investment techniques versus conventional financial analysis. Implicitly they also compared the different fund managers. Analogously to the various social screens applied across different funds, active management will not be similar for each fund either. Therefore, any divergence in performance can be largely dependent on the discretionary choices of a single fund manager (Kempf & Osthoff, 2007).

3.4 Sin Stock Studies on Performance of Responsible Investment

The avoidance of sin stocks is not always the same as responsible investing. Negative screening should be considered as a sub strategy. Tobacco, alcohol, and gambling are the most often described controversial business areas and sometime referred to as the 'Triumvirate of Sin'. These three are considered especially heinous because of their addictive nature. Furthermore, they have limited substitutes. Therefore, many countries charge excise taxation to make them more costly (Hong & Kacperczyk, 2009; Salaber, 2007).

Studying the performance effects of negative, sin stock screening means a deviation from traditional finance theory. Institutional investors have a fiduciary obligation to make money, and according to conventional pricing models, individual moral beliefs should not play a role in stock picking. With the capital asset pricing model (CAPM), all that should count is the market risk premium (Fabozzi, Ma & Oliphant, 2008). Yet, norms-constrained investors like pension funds are still willing to pay a financial cost to withhold from certain stocks. This sort of behaviour means that the CAPM does not always hold, which can be the case for segmented markets or neglected stocks (Hong & Kacperczyk, 2009). In such instances, Merton (1987) argues that not only sensitivity to market risk and return – beta – matters, but that idiosyncratic risk also plays a role.

Figure 3. Effect of shunning sin stocks.



Source: "Effect of Investment Action", Statman (2000; p.36) Socially Responsible Mutual Funds

If large groups of investors neglect a certain stock, its price can fall below fundamental value, leading to higher returns. Figure 3 demonstrates that if the market withdraws or withholds capital from perceived controversial companies, the capital supply curve of such a company shifts from S_1 to S_2. When the demand curve is not perfectly elastic to supply, the y-axis shows that the expected return by investors and thereby the cost of capital for a company is raised. Prerequisite for establishing a change in the cost of capital and therefore the irresponsible company's behaviour is the absence of substitute capital. When responsible investors shun a certain stock, but conventional investors provide the desired capital anyhow, the cost of capital will not be affected.

Hong and Kacperczyk (2009) showed that the neglect of stocks by large institutional investors had a profound effect on the prices of sin stocks, ranging from 15-20 percent. Other studies showed varying results, which can be attributed to how confined the definition of 'sin' is. For instance, besides the triumvirate of sin other areas like weapons production can also be considered as being irresponsible. Moreover, in some European countries (e.g. Luxembourg) weapons like cluster ammunition are banned by law (Ethix SRI Advisors, 2011), whereas U.S. investors do not all consider weapon producers as controversial. This different interpretation of what is controversial and the consequent breadth of the definition has caused the evidence of studies to be either statistically significant, or not (Statman & Glushkov, 2009).

Salaber (2007) found that the height of the excess return on sin stocks is locally determined by legal and cultural specificities. For example, protestant countries were found to be more 'sin averse' than Catholic nations, requiring a significant premium. Moreover, not only religion influences the excess returns, also the level of excise taxation and the degree of litigation risk play a significant role. Higher taxation and litigation risks are found to lead to higher expected returns. Litigation risk and 'headline risk' (i.e. negative news) result in a permanent discount in the stock price. The perceived costs of lawsuits and settlements are immediately incorporated in the market price of the stock. Disclosure or bad reporting were not related to the higher returns, since Kim and Venkatachalam (2011) found that controversial companies often have better reporting systems to account for the extra litigation risk they face.

3.5 Portfolio-based Studies on Performance of Responsible Investment

The downsides of mutual fund studies on the performance of responsible investing resulted in a stream of research based on self-constructed portfolios. To be more independent of the discretionary choices of the investment fund manager and for specifically analysing the effects of certain screening policies, scholars stopped using portfolios of mutual funds and constructed their own stock portfolios. This enabled them to better separate the sources and explanations of the

under- and outperformance of responsible investment, making leeway for the explanation of any abnormal returns due to screening (Derwall et al.,2011; Kempf & Osthoff, 2007).

Although investment managers usually install a multitude of screens, the early studies based on self-constructed portfolios predominantly focused on environmental matters to deem which stocks were responsible. Yet, for studies to stay aligned with the real-world investment practice, a broader set of criteria was needed to determine which companies should be listed as responsible. Therefore, later studies combined several ESG factors when establishing a portfolio of responsible stocks. These factors include both involvement and performance of companies concerning alcohol, community relations, diversity, employee satisfaction, environment, military, tobacco and so on.

Again, the academic evidence on the financial influence of these factors is mixed. Diltz (1995a) showed that a screening strategy based on environmental performance and exclusion of military involvement delivers significant positive returns. Yet, other scholars did not find any statistically significant performance differentials after incorporating non-financial issues (Guerard, 1997). It might be that certain ESG factors influence financial performance more than others do. Edmans (2011) focused solely on employee satisfaction and found a positive correlation between this factor and financial performance. Furthermore, he showed that achieving abnormal returns based on the employee satisfaction factor is possible, thereby concluding that the stock market does not fully incorporate this non-financial information in prices.

The most common way in which these studies try to establish whether ESG screening strategies deliver abnormal returns involves the comparison of different portfolios. Often a distinction is made between low- and high-rated stocks on predetermined ESG criteria. Consequently, these respective stocks are structured in two different portfolios. The returns of these individual portfolios are then compared to performance benchmark models like the Capital Asset Pricing Model (CAPM)[7]. Furthermore, these results are often supplemented by employing a long-short strategy, buying the high-rated portfolio and selling the low-rated portfolio, and looking if this delivers any significant abnormal returns (Kempf & Osthoff, 2007; Statman & Glushkov, 2009; Derwall et al., 2011).

The two most recent and most elaborate studies using self-constructed portfolios to examine the performance of responsible investment in the U.S., used the sustainability data of Kinder, Domini and Lydenberg (KLD). Stocks of companies listed in the U.S. are rated by KLD on several ESG factors. Based on these ratings Kempf and Osthoff (2007) and Statman and Glushkov (2009) subdivided stocks into portfolios to test the effects of negative, positive and best-in-class

[7] Different performance benchmark models will be reviewed in chapter 4.

screening policies. The assumed underperformance of negative screening was tested by creating a low-rated stock portfolio of companies involved in at least one controversial business area. In this case, perceived controversial businesses are determined according to the exclusionary criteria provided by KLD: alcohol, tobacco, gambling, military, firearms, and nuclear power. Note this is a broader definition of controversial as compared to the triumvirate of sin used by Hong and Kacperczyk (2009). A similar approach was used for positive screening, in which the low-rated portfolio comprised stocks with low KLD ratings and the high-rated portfolio contained high ranked stocks.

Kempf and Osthoff (2007) reviewed all constituents of the S&P500 and the DSI400 across a time span ranging from 1992 to 2004. Their results showed that a basic long-short strategy based on ESG factors could deliver abnormal returns as measured by the Carhart (1997) four-factor model. Positive returns were found by employing the positive and best-in-class screening strategies. Abnormal returns were strongest with the best-in-class screening method and only when the extreme ratings were used, maintaining very small cut-off points of merely including the top- and bottom 5 percent of the rated companies. Consistent with the findings on sin stocks by Hong and Kacperczyk (2009), a negative screening policy delivered negative returns when buying a portfolio of non-controversial companies and shorting a portfolio of sin stocks (Kempf & Osthoff, 2007).

Statman and Glushkov (2009) also investigated the returns on responsible investment in the U.S during an extended period, from 1992-2007. They tested three – earlier discussed – hypotheses: (1) “Doing Good While Doing Well”, (2) “Doing Good, but Not Well”, (3) “No Effect”. Their conclusion is that the first hypothesis is possible with a positive or best-in-class screening strategy. Yet, in case of a negative screening policy, the second claim also holds. Moreover, they assert that when screening strategies are combined, the positive and negative effects can offset each other resulting in the third hypothesis. Therefore the true ethical investor, heavily using negative screens, might not benefit from responsible investing, since they might find the most promising method (best-in-class screening, without exclusion) not responsible at all.

The abnormal returns of positive screening are not only witnessed in the U.S. The Sustainable Asset Management Group (SAM, 2011) studies the financial performance of the 2,000 largest companies worldwide, using their own proprietary corporate sustainability data. Their final sample covers the years from 2001 to 2010, including 465 companies per year. The results show that a long position in the portfolio of ‘sustainability leaders’ and a short position in ‘sustainability laggards’ delivered an annual outperformance of 3.68 percent, with a T-statistic of 2.25. By exploiting these less researched non-financial factors, SAM believes it is possible to create long-term value, with robust results in bull and bear markets. Defined as an effective “all-weather

approach" they claim that combining financial and extra-financial factors leads to optimal portfolios with improved risk characteristics (p.3).

With multiple studies confirming the outperformance of companies scoring high on ESG characteristics, it is plausible that incorporating sustainability ratings adds value-relevant information to the investment process. Yet, the question remains what causes the outperformance. The already introduced 'errors-in-expectations' hypothesis claims it is caused by markets mispricing the impact of good ESG performance. This mispricing claim is true if ESG factors actually influence a company's cash flow and synchronously not enough information about this effect reaches the market, resulting in underestimated stock prices. Eventually, when the benefits of ESG efforts materialize and market participants do recognise the upside, stock prices adjust accordingly and abnormal returns can be observed.

Besides this mispricing theory, some suggest that positive alpha does not exist and that it simply signifies 'undefined' beta. That is, alpha values are caused by a yet to be identified risk factor explaining the abnormal returns. Mănescu (2011) studied the potential existence of a non-sustainability risk factor. Yet, following that study, positive outperformance derived from incorporating the KLD ESG indicators is not due to a missing risk factor. The study concludes that some ESG elements have an impact on value. However, since this value-relevant information is not always efficiently captured in stock prices, the outperformance can be attributed to mispricing (Derwall et al., 2011; Kempf & Osthoff, 2007; Mănescu, 2011).

Finally, research arguing that ESG incorporation can lead to enhanced portfolio return has been published for quite some time now. Knowing that responsible investment has more and more become a mainstream investment approach, it could be argued that the gross majority of investment managers has started to pay attention to non-financial information. This is an assertion that is evidenced by the growing number of signatories to the United Nations' Principles for Responsible Investment. With investment managers incorporating extra-financial information as quickly and thoroughly as possible, it could very well be that the outperformance due to false expectations becomes increasingly short-lived (UN PRI, 2012; US SIF, 2010).

In the upcoming chapters, the performance of the described screening strategies will be empirically tested. The focus of this thesis lies on a recent data set to check whether the market integrates the benefits of ESG information more efficiently or that positive abnormal returns based on ESG information can still be achieved.

4. Methodology and Empirical Implementation

4.1 Research Philosophy

Business research often distinguishes between two research philosophies: positivism and phenomenology. The former applies reductionist methods of observing, measuring, and describing existing events in order to find relationships. The latter is more concerned with the meaning of phenomena; questioning and taking into account feelings and beliefs. Although both streams are not completely different and separable, empirical research is often conducted with a positivist research philosophy. For empirical research to reach generalizable results, it is important that the findings can be replicated. Therefore, experiments and tests have to be controlled and the results have to be argued for validity of findings and randomness of errors (Remenyi, Williams, Money & Swartz, 1998).

According to Remenyi et al. (1998) this "bias for empiricism" demands evidence to reach valid conclusions, but it also assumes theoretical knowledge and some preposition on the research topic (p.31). "Theory cannot be generated without data and data cannot be collected without a theoretical framework" (p.32). Therefore, this thesis used a hypothetic-deductive method, combining a theoretical framework and hypothesis testing (Sekaran, 2003). The theoretical review in the previous chapter described the factors of influence on the performance of responsible investment. In the next section, hypotheses are constructed to empirically test the performance of different screening methods on historical data.

4.2 Hypothesis Development

The empirical part of this thesis will test several screening methods to establish whether portfolios based on ESG ratings can still deliver abnormal returns. With all previous studies on the performance of responsible investment combined, it is further hypothesised that different types of responsible investors coexist. The first group of 'values-based investors' would be best served by negative screens, excluding companies that operate in according to them sinful or controversial business areas. These investors are primarily motivated to invest in line with personal values and are concerned with the societal impact, attaching less value to the financial returns. Specific studies on sin stock returns have shown the existence of abnormal returns for controversial stocks. The first hypothesis will test whether this 'shunned stock' effect is also present within the data set and time range used in this thesis:

H1: A low-rated portfolio with stocks of companies involved in at least one controversial business area will outperform a high-rated portfolio consisting of all other stocks.

The other group of 'profit-seeking' investors thinks that a good ESG record can serve as a proxy for financial outperformance. These investors are best served by using positive screens in order to pick the stocks of companies that perform best on ESG matters. One of the assumptions for this to be possible is the fact that the market must misprice the value-adding potential of sustainability efforts. Yet, knowing that responsible investing has grown rapidly and market interest has caught on, it is plausible that such mispricing will become increasingly short-lived now responsible investment comes to be a mainstream investment method. Therefore, the second hypothesis will check the up-to-date validity of such an outperforming 'errors-in-expectations' effect:

H2: A high-rated portfolio with stocks of companies with high ESG ratings will outperform a low-rated portfolio with stocks of companies with low ESG ratings.

Finally, the theoretical framework showed that a specific positive screening method, the best-in-class approach, delivered even higher returns compared to conventional positive screening. The third hypothesis will investigate the performance of this best-in-class strategy:

H3: A high-rated portfolio with stocks of companies with high best-in-class ESG ratings will outperform a low-rated portfolio with stocks of companies with low best-in-class ESG ratings.

4.3 Data Description

Data set of ESG ratings

Essential to the portfolio comparison used in this empirical study is research on ESG criteria, in order to determine which stocks can be depicted as good or bad performers according to ESG factors. Nowadays, plenty of ESG research providers exist. The most prominent names include ASSET4, EIRIS, Innovest, MSCI ESG Research, SiRi Company, Sustainable Asset Management Group (SAM) and Trucost. Yet, most of these providers have a proprietary rating methodology and when it comes to sharing their databases for research, the options become sparse (Mănescu, 2011; Statman & Glushkov, 2009). The only freely accessible ESG data are the ratings provided by KLD Research & Analytics, the same source as used in many of the previous studies on the performance of responsibly invested portfolios. This however constrains the universe to only the largest U.S. listed companies and blocks any initial ambitions to provide a European equivalent to the Kempf and Osthoff (2007) and Statman and Glushkov (2009) studies.

The KLD data has longest record of ESG data and is favoured in many studies, offering a reliable source of ESG ratings for this thesis. Moreover, compared to previous studies the universe of rated companies has been extended and the most recent years have not yet been included in previous studies, giving a new unique angle to the analysis that will follow in the upcoming chapters.

The KLD dataset is retrieved from the Wharton Research Data Services (WRDS). Historically, KLD Research & Analytics was the company that provided the ratings. However, in 2009 their methodology and activities were sold to the RiskMetrics Group, which was consequently acquired by MSCI the following year. Therefore, the official data provider now is MSCI ESG Research, who yearly publishes the spreadsheet containing ESG ratings. This spreadsheet, the MSCI ESG Statistical Tool for Analysing Trends in Social and Environmental Performance (STATS), contains the following information:

(1) Company identifiers (e.g. name, ticker, CUSIP);
(2) 50 ESG strength and concern indicators across seven qualitative categories:
    a. *Community;*
    b. *Corporate Governance;*
    c. *Diversity;*
    d. *Environment;*
    e. *Employee Relations;*
    f. *Human Rights;*
    g. *Product;*
(3) Six involvement measures for excluding controversial businesses:
    a. *Alcohol;*
    b. *Gambling;*
    c. *Firearms;*
    e. *Military;*
    f. *Nuclear Power;*
    g. *Tobacco.*

The seven qualitative categories are subdivided in both strengths and concerns, but the six exclusionary measures only list a negative indicator when companies are involved in the corresponding controversial business areas. Please refer to table I in the appendix, which shows all strengths and concerns used in the analysis for this thesis, for an overview of the subdivision of qualitative indicators. The data in the STATS spreadsheet is displayed in a binary form. When a strength, concern, or involvement indicator is found at a certain company, this is represented by a '1'. A '0' indicates that that specific indicator is not present at the company.

As shown in table 4a the stock universe that is covered by the KLD/MSCI data has been extended over the years. Previous research (Kempf & Osthoff, 2007; Statman & Glushkov, 2009) focused on the companies listed in either the S&P500 or the DSI400, but for this thesis the entire data set of all listed companies in the Russell3000 is used, resulting in nine years of ESG data

(2003-2011) on approximately 3,000 companies (KLD Research & Analytics, 2006; MSCI ESG Research, 2011).

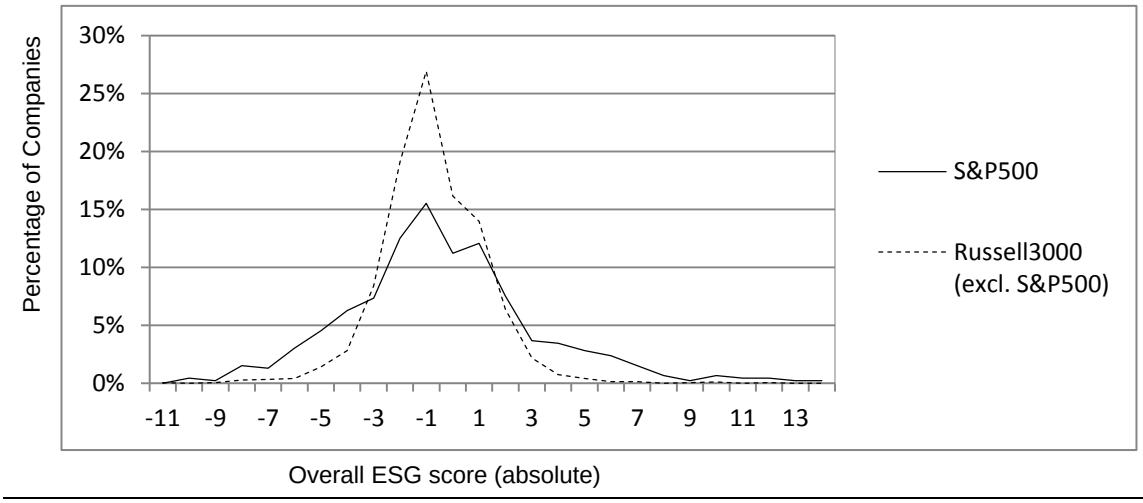
Table 4a. KLD/MSCI ESG STATS coverage

Coverage Universe	1991-2000	2001-2002	2003-present
500 Largest US Companies	X	X	X
MSCI KLD 400 Social Index	X	X	X
1,000 Largest US Companies		X	X
3,000 Largest US Companies			X
Approximate Total Number of Companies Covered	650	1100	3000

Source: MSCI ESG Research (2011)

The distributions of the cumulative qualitative ESG scores of companies in the S&P500 and the Russell3000 are shown in figure 4a. The mean scores are respectively -0.36 and -0.75, but the most significant difference is found in the fourth moment of the distribution. The S&P500 data displays low Kurtosis compared to the Russell3000 curve. A high-peaked distribution indicates that a lot of the variability is due to extreme values: fat tails. In the performance measurement, this will be taken into consideration by altering the cut-off points that determine which fraction of stocks is included in the low- and high-rated portfolios. In earlier studies, these have varied between 5 to 50 percent. Compared to the practices in S&P500 studies a smaller cut-off for the Russell3000 universe might be desired to capture large enough ESG differentials (Bodie et al., 2009).

Figure 4a. Distribution of cumulative qualitative scores of companies in the S&P500 and the Russell3000. The cumulative scores are calculated by summing all strengths and subtracting all concerns of the seven qualitative indicators. S&P500 companies also part of the Russell3000 are excluded to prevent overlap.



Data: MSCI ESG STATS (2012)
Adapted from: Statman & Glushkov (2009)

As previously noted, screening choices are always grey. The overall ESG scores are indicative for this fact. Table 4b shows the composition of the scores per qualitative indicator for some randomly drawn companies from the data set. A high-ranking company does not necessarily perform well on all indicators, nor is a low-ranked company the worst performer on all. For example, although Wal-Mart displays the worst scores on the 'employee relations' and 'product' indicators, it is surpassed by the Exxon Mobil Corporation when it comes to the lowest overall ESG score (Statman & Glushkov, 2009).

Table 4b. Cumulative qualitative scores of companies in the Russell3000, 31 December 2009. The companies in this overview exemplify how the absolute overall ESG scores can be constructed based on the individual qualitative indicators.

Social Characteristic	Exxon Mobil Corp.	Wal-Mart Stores Inc.	Goodyear Tire & Rubber Co.	Goldman Sachs Group Inc.	Delta Air Lines Inc.	Procter & Gamble Co.	General Mills Inc.	Intel Corp.
Community	-2	0	-1	0	0	0	1	1
Corporate Governance	-2	-3	-1	-2	0	0	0	1
Diversity	2	3	-1	3	1	5	4	5
Employee Relations	0	-3	-2	2	-1	1	2	5
Environment	-4	-1	-1	0	0	0	1	2
Human Rights	-2	-1	0	-1	0	0	0	0
Product	-2	-4	-2	-3	0	0	2	0
Overall score	**-10**	**-9**	**-8**	**-1**	**0**	**6**	**10**	**14**

Data: KLD/MSCI ESG STATS (2012)
Adapted from: Statman & Glushkov (2009)

The relationship between the seven qualitative indicators is shown in table 4c. The cross-sectional correlations show that no single qualitative criterion is indicative for the overall ESG score, but all are positively related within a range of 0.38 to 0.53. Amongst the seven individual indicators, the correlations are also not that high, the correlation between 'community' and 'environment' being the highest with 0.23 (Kempf & Osthoff, 2007).

Table 4c. Correlation matrix of the seven qualitative screens in the MSCI ESG STATS, 2003-2010

	Overall ESG score	COM.	CORP.GOV.	DIV.	EM.REL.	ENV	HUM.RIG.	PRO.
Overall ESG score	***1***							
Community score	***0.45***	1						
Corporate Governance score	***0.38***	0.01	1					
Diversity score	***0.53***	0.10	-0.23	1				
Employee Relations score	***0.40***	0.09	0.04	0.02	1			
Environment score	***0.50***	0.23	0.11	0.11	0.08	1		
Human Rights score	***0.26***	0.05	0.12	-0.07	0.07	0.13	1	
Product score	***0.36***	0.04	0.20	-0.17	0.08	0.08	0.14	1

Data: KLD/MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007)

Up until now the scores have been calculated as a simple sum of strengths and concerns, yet this can give a skewed image since not all qualitative indicators share and equal number of strengths and concerns. Moreover, the composition of these sub criteria has changed over the years, also inhibiting a year-by-year comparison based on absolute scores. Therefore, some kind of aggregation method has to be used to normalize the scores and make them comparable across the different ESG indicators. Kempf and Osthoff (2007) took the binary complements for all the concerns, so that the lack of a concern would be shown as a strength. For instance, normally the absence of any investment controversies would be indicated as a '0' for the corresponding community concern criterion. However, by taking binary complements, the data set shows a '1' if a company lacks a concern, turning it into a strength criterion. As shown in formula 1, the sum of the strengths and binary complements of the concerns, relative to the total number of strengths and concerns, delivers a comparable ESG score across indicators and years (Kempf & Osthoff, 2007).

Kempf & Osthoff (2007): $$ESG\ {}_t^j = \frac{\sum_{s=I}^{u_t^j} strenght_s^j - \sum_{r=I}^{k_t^j} (!weakness_r^j)}{u_t^j + k_t^j} \quad (1)$$

With:

$ESG\ {}_t^j$: $ESG\ characteristic\ j, year\ t$;:

$strenght_s^j$: $strength\ indicator = \begin{cases} 1, \text{if firm meets strength s on characteristic j;} \\ 0, \text{otherwise;} \end{cases}$

u_t^j : $number\ of\ strengths\ for\ characteristic\ j, year\ t$;

$weakness_r^j$: $concern\ indicator\ = \begin{cases} 1, \text{if firm meets weakness r on characteristic j;} \\ 0, \text{otherwise;} \end{cases}$

k_t^j : $number\ of\ concerns\ for\ characteristic\ j, year\ t$.

Yet a recent study by Mănescu (2011) criticises this method, since the absence of a concern should not immediately be interpreted as a strength. Environmental criteria like 'Land Use & Biodiversity' are less relevant for a consultancy firm than for an industrial company. Nevertheless, with the method displayed in formula 1 scores of companies that are not involved in any kind of soil operations would automatically receive an enhanced environment score. If this happens in multiple indicators, it will ultimately lead to inflated ESG scores. Mănescu opted for a new method, summing the strengths and concerns relative to the total number of strength and concern indicators used. This relative method will also be used for the data analysis in this thesis. Formula 2 shows how the ESG scores are calculated per qualitative indicator. Consequently, to establish an overall annual ESG score for a single company, the simple average over the seven qualitative indicators can be calculated; as shown in formula 3.

Mănescu (2011): $$ESG\ {}_t^j = \frac{\sum_{s=1}^{u_t^j} strenght_s^j}{u_t^j} + \frac{\sum_{r=1}^{k_t^j} weakness_r^j}{k_t^j} \quad (2)$$

Mănescu (2011): $$\sum_{j=1}^{7} \frac{ESG_t^j}{7} \quad (3)$$

Both the individual scores per indicator and the average overall ESG score will be used in the analysis. The individual scores will be used to test whether some ESG indicators offer a better perspective for financial outperformance. Paragraph 4.4 will highlight how these portfolios will be constructed, but first the financial data used will be reviewed.

Data set of financial returns

The financial returns of the all companies rated in the KLD/MSCI ESG STATS need to be included in the analysis as well. For this purpose, the Center for Research in Security Prices (CRSP) database is used, which is also accessed via WRDS. The CRSP U.S. Stock Database offers a broad-ranging collection of return calculations for NYSE, AMEX and NASDAQ stock markets and is advertised as having the "most accurate total return calculations" (WRDS, 2012). The specific data set used is the annually updated CRSP Monthly Stock file. The cross-sectional sample ranges from January 2004 to December 2011. The data gathered comprises identifying information (company name, ticker, CUSIP, SIC-code) and the returns, the latter being the monthly holding period returns (Sekaran, 2003).

The ESG ratings data provided by KLD/MSCI is free of survivorship bias. Yet, the final samples used for the portfolio analysis in this thesis only incorporate companies with a complete overview of CRSP monthly holding period returns; incomplete being less than twelve holding period returns per annum. This means that companies whose stocks got delisted during a certain year are not included in portfolios of that respective year, which introduces a survivorship bias in those samples. This might result in the holding period return data being skewed upwards, since the delisted stocks are likely to have underperformed and are only partially represented in the annual CRSP return data due to take-overs or delisting. There are two juxtaposed academic views on survivorship bias, one leg (Brown, Goetzman, Ibbotson & Ross, 1992) arguing for overblown performance and the other (Grinblatt & Titman, 1992) claiming reversal of performance. It is understood that survivorship bias is an important topic, but given the ambiguous influence on performance, the deletion of companies with an incomplete record in this study is not further related to any performance measurement (Chegut, Schenk & Scholtens, 2011; Hallahan & Faff, 2010). The eventual analysis will mainly look at the performance of a long-short strategy

comparing any abnormal returns of both the low- and high-rated portfolios. It is assumed that both will be equally exposed to any effects of this bias.

Additionally, the KLD/MSCI data set contained companies with absolute overall ESG scores of zero. Sometimes this is due to qualitative indicators cancelling each other out, but some other companies did not have any listed strengths or concerns at all. It is debatable whether these latter companies have been researched and rated by KLD/MSCI. Similar to Statman and Glushkov (2009) this thesis will omit these "no-indicators-zero" (NIZ) companies from the final sample (p.37). A summary of the data retrieved for the empirical analysis can be found in table 4d. It shows the number of companies in the KLD/MSCI ESG STATS yearly spreadsheet, the number of deleted companies due to NIZ or incomplete returns series, and the eventual CRSP U.S. Monthly Stock file observations.

Table 4d. Overview of the final yearly data sets used in the analysis. Deleted companies are ones with incomplete return series or those without a rating (no-indicators-zero). The cut-off points show how many companies will be incorporated in both the high- and low-rated portfolios. The last two columns display the number of companies that were involved in at least one perceived controversial business area.

	Number of Companies					Accepted versus Shunned	
Year financial data	Companies rated in KLD dataset	Deleted companies	Remaining monthly observations	Cut-off point 33 percent (no. of stocks)	Cut-off point 5 percent (no. of stocks)	Accepted companies	Shunned companies
2004	2311	406	22860	635	95	1765	140
2005	2706	435	27252	757	113	2119	152
2006	2728	432	27552	765	114	2127	169
2007	2778	472	27672	768	115	2137	169
2008	2737	372	28380	788	118	2189	176
2009	2723	283	29280	813	121	2252	188
2010	2664	271	28716	797	119	2207	186
2011	2771	189	30984	860	129	2425	157

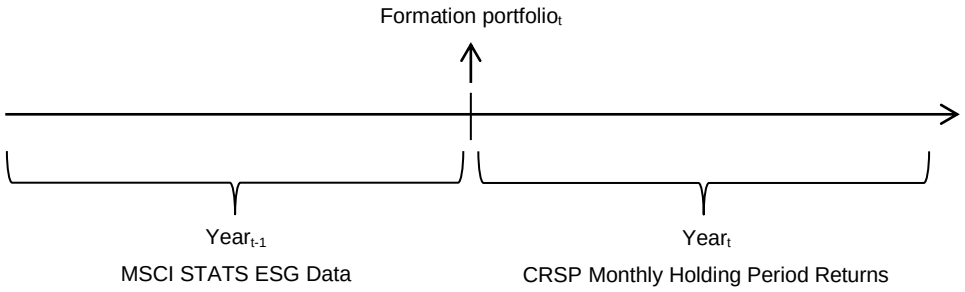
Data: MSCI ESG STATS (2012), CRSP (2012)

4.4 Portfolio Formation

The years over which the data is collected diverge. The KLD/MSCI ESG STATS data is collected from 2003 to 2010, but the financial returns are collected from 2004 to 2011. Since the ESG ratings are published at years' end, they serve as the input for the creation of portfolios for the consequent year. Figure 4b shows how the rating of year t-1 serves as the input for the portfolios of year t. The portfolios are annually reconstructed at the beginning of year t, adjusting the portfolio constituents according to any changes in the rating over year t-1.

Because the first ESG ratings for the entire Russell3000 universe date back to 2003, it is only possible to start constructing portfolios from 2004 onwards. Similarly, although the 2011 ESG ratings are already available, no full year return series over 2012 are recorded; resulting in the time span from 2004 to 2011.

Figure 4b. Graphical overview of the portfolio formation. Portfolios for year t are constructed with ESG ratings data published over year t-1.



From the literature review, it has become apparent that there is not yet a universal best practice method for setting up screening strategies. The performance studies make many arbitrary choices, for instance in choosing which fraction of top- and bottom-performing companies should be included in a portfolio. Diltz (1995b) even concludes that social screening is “more an art than a science” (p. 70). Therefore, this thesis will carefully describe how the portfolios used in the analysis are formed.

Based on the year t-1 ratings, this thesis constructs several low- and high-ranked portfolios. The first hypothesis, examining the ‘shunned stock’ effect, focuses on companies that are involved in perceived controversial business areas. The KLD/MSCI ESG STATS data shows involvement on six exclusionary indicators. If a company is involved with alcohol, firearms, gambling, military, nuclear power, or tobacco in a given year, it is classified as a sin stock and placed in the low-rated portfolio of shunned stocks. All other – accepted – stocks are placed in the high-ranked portfolio.

Next, the second hypothesis investigates the possible outperformance of positive screening due to an ‘errors-in-expectations’ effect. For this multiple ESG scores will be examined. First, each of the seven qualitative indicators will be tested individually. Stocks are ranked per year from high to low according to their relative ESG score as calculated by formula 2. Consequently, for each year the top and bottom 33 percent are positioned into respectively the high- and low-rated portfolios. In comparison to previous studies, this thesis also includes the ‘Corporate Governance’ indicator. Corporate Governance is often excluded from the analysis, since it was only added as such in 2002. Before that time, it was labelled ‘Other’. Studies covering the pre- and post-2002 period chose not to incorporate this mixed indicator because it is unclear whether the definitions can be aligned. This problem is not present in the time span used for this thesis, since it starts after 2002 (Kempf & Osthoff, 2007; Statman & Glushkov, 2009; Mănescu, 2011). Lastly, yearly top and bottom portfolios are also created for the overall ESG score as calculated by formula 3; again taking 33 percent as the cut-off point.

For the final hypothesis on best-in-class screening, also the yearly overall ESG scores are used. However, for this hypothesis they are balanced across ten industries. Based on the industry definitions provided by the Kenneth R. French Data Library (2012), the stocks are partitioned into ten industries according to their four-digit Standard Industry Classification (SIC) code[8]. Within each industry the top and bottom 33 percent are added to respectively the high- and low-ranked portfolios.

4.5 Performance Measurement

To couple the stocks in the portfolios with their corresponding monthly holding period returns the eight-digit CUSIP codes are used, which proved to be a more reliable indicator than the company tickers. Per high- and low-rated portfolio, this has delivered 8 years of 12 holding period returns observations, totalling to a series of 96 returns for each portfolio over the entire cross-section.

Naturally, the monthly *portfolio* returns depend on the holding period returns of the individual stocks in the portfolios. In practice, portfolio returns are frequently value weighted, but in this thesis, an equally weighted method is used for the computation of the monthly portfolio returns. Earlier studies have often used both equally and value weighted schemes to calculate the returns of a portfolio of stocks. Based on the findings of these studies the equally weighted approach is chosen. The excess returns of a value-weighted portfolio have a tendency to be lower compared to an equally weighted scheme. Moreover, when looking at the *p*-values of the results, the statistical significance of value-weighted returns was generally lower (Statman & Glushkov, 2009).

Capital Asset Pricing Model

The most straightforward methods for risk-adjusted performance evaluation are closely related to the Capital Asset Pricing Model (CAPM). In the wake of the development of modern portfolio theory and the 'mean-variance' efficient portfolio by Markowitz (1952; 1959) the CAPM was concurrently introduced by Sharpe (1964), Lintner (1965) and Mossin (1966). The model yields an expected return for a security or portfolio by relating the individual risk premium on an asset to the market risk premium; the market risk premium is the return over the risk-free rate for holding the market portfolio and thus a compensation for non-diversifiable risk. The well-known beta (β) coefficient measures the co-movement of an asset with the market. It can be captured by the Ordinary Least Squares (OLS) "estimator of the slope coefficient in the excess-return market model", which is the beta coefficient in regression equation 4 (Bodie et al., 2009; Campbell, Lo & MacKinlay, 1997; p.183).

[8] Please refer to table II in the appendix for the Kenneth R. French Data Library industry definitions.

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it} \quad (4)$$

α_i $Jensen's\ alpha;$
R_{it} $Return\ on\ portfolio\ i\ in\ month\ t;$
R_{ft} $Return\ on\ one\ month\ Treasury\ bill\ in\ month\ t;$
R_{mt} $Return\ on\ equity\ benchmark\ in\ month\ t;$
ε_{it} $Error\ term.$

(Bauer, et al., 2005; p. 1758)

Risk-adjusted performance measures based on the CAPM are the Sharpe measure, the Treynor measure, Jensen's measure, and the information ratio. Jensen's measure, the intercept of the CAPM, also referred to as alpha (α), measures the under- or outperformance of a portfolio relative to the market. In most studies that compare portfolio returns, it is the performance measure of choice. In the analysis for this thesis, the Jensen's measure will also be employed. As shown by the regression model in formula 4, alpha states the return in excess of the expected return stipulated by the CAPM, "given the portfolio's beta and the average market return" (Bodie et al., 2009; p.826).

Although Jensen's measure is widely applied in performance studies on responsible investment, it is not without critique. The use of the CAPM model does not withstand all empirical tests and can result in pricing errors. Therefore when interpreting the Jensen's measure, an alpha value can signal either outperformance or an improper benchmark (Bauer et al., 2005; Bodie et al., 2009; Statman & Glushkov, 2009).

Multi-factor models

If the CAPM causes a mispricing that can be exploited by arbitrage, it means that the relationship between risk and return is not fully explained by the model. Building on the thought that systematic risk can be multi-layered; this thesis will try to overcome these pitfalls of the single-index CAPM by also including multi-factor models in the tests on portfolio performance. The most commonly used multi-factor models in performance studies on responsible investment are the Fama-French (1993) three-factor model and the Carhart (1997) four-factor model. These models provide guidance on which factors, or sources of risk, contribute to the risk premium and are not covered by the CAPM. Although these multi-factor models are also not free from discussion, they have become the standard for performance measurement (Bodie et al.,2009; Statman & Glushkov, 2009).

The Fama-French (1993) three-factor model expands the CAPM with factors correcting for "firm size and book-to-market ratio" (Bodie et al., 2009; p. 423). The small-minus-big (SMB) factor

accounts for the diverging risk characteristics of small versus large companies, as measured by their market capitalization. Historically, small firms have proven to yield higher than predicted returns; a peculiarity adjusted for with the SMB risk factor. Likewise, the high-minus-low (HML) factor corrects for the historic outperformance of firms “with high ratios of book equity to market equity”, also popularly called ‘value’ companies (p. 423).

The Carhart (1997) four-factor model covers another anomaly in market prices. The added momentum (MOM) factor corrects for the fact that prices have a tendency to display path dependency in their returns. ‘Winners’, stocks with positive returns over the previous periods, are likely to keep performing well and vice versa for ‘losers’. The MOM factor includes the return difference that is observed between portfolios with strong and weak past performance over the forgone year.

Altogether, since these three benchmark models incorporate different risk factors, the resulting estimations for the alpha values will vary as well, dependent on the respective tilts towards each risk factor. Formula 5 shows the regression models for both the Fama-French (1993) three-factor model and the Carhart (1997) four-factor model. When omitting the $\beta_{3i}MOM_t$ term, the three-factor alpha is estimated; otherwise, the model will deliver a four-factor alpha (Bauer, et al., 2005; Bodie et al., 2009; Statman & Glushkov, 2009).

$$R_{it} - R_{ft} = \alpha_i + \beta_{0i}\left(R_{mt} - R_{ft}\right) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{3i}MOM_t + \varepsilon_{it} \quad (5)$$

MKT_t $\left(R_{mt} - R_{ft}\right)$, $excess\ return\ on\ equity\ benchmark\ in\ month\ t$;
SMB_t "$the\ difference\ in\ return\ between\ a\ small\ cap\ portfolio\ and\ a\ large\ cap\ portfolio\ at\ time\ t$;”
HML_t "$the\ difference\ in\ return\ at\ time\ t\ between\ a\ portfolio\ containing\ 'value'\ stocks\ (with\ a\ high\ book-to-market\ ratio)\ and\ one\ consisting\ of\ 'growth'\ stocks\ (with\ a\ low\ book-to-market\ ratio)$;"
MOM_t “$the\ difference\ between\ a\ portfolio\ of\ past\ 12\ months\ winners\ and\ a\ portfolio\ of\ past\ 12\ month\ losers\ at\ time\ t$."

(Bauer, et al., 2005; p. 1760-1761)

This thesis does not want the results to be dependent on the flaws of one particular model. Therefore, the excess returns of high-, low-rated and long-short portfolios are examined according to all three above introduced benchmark models. Given the importance of choosing a proper benchmark, the CAPM is extended with both the Fama-French and Momentum factors, to overcome that outperformance is resulting from any unexposed risk factors (Edmans, 2011; Statman & Glushkov, 2009).

For the three benchmark models, the regression inputs are taken from the Kenneth R. French Data Library (2012), which provides research factors that are widely used in academic research. The data library creates the excess return on the market by using the CRSP value weighted index, which includes all NYSE, AMEX, and NASDAQ stocks. For the risk-free rate, the one-month Treasury bill rate is used. The SMB, HML, and MOM factors are all calculated by segmenting the NYSE, AMEX, and NASDAQ stocks in portfolios. For the SMB and HML factors six portfolios based on the market capitalization and book-to-market ratio are created[9]. The MOM factor is based on portfolios subdivided by size and the prior (2-12 monthly) returns[10].

[9] $SMB = \frac{1}{3}(Small\ Size\ Value + Small\ Size\ Neutral + Small\ Size\ Growth) - \frac{1}{3}(Big\ Size\ Value + Big\ Size\ Neutral + Big\ Size\ Growth)$

$HML = \frac{1}{2}(Small\ Size\ Value + Big\ Size\ Value) - \frac{1}{2}(Small\ Size\ Growth + Big\ Size\ Growth)$

[10] $MOM = \frac{1}{2}(Small\ Size\ High\ Return + Big\ Size\ High\ Return) - \frac{1}{2}(Small\ Size\ Low\ Return + Big\ Size\ Low\ Return)$

5. Data Analysis and Empirical Findings

5.1 Hypotheses Testing

The hypotheses that were introduced in section 4.2 will be tested in this chapter. Jensen's measure will be used to test whether portfolios built on the basis of ESG factors can yield abnormal returns. Besides the standard single-index alpha, the analysis is extended by also employing the three- and four-factor alphas of respectively Fama-French (1993) and Carhart (1997). The alpha values are calculated with the regression models introduced in chapter four, stated by formulas 4 and 5. The input for the dependent variable is the monthly excess holding period return on the portfolio ($R_{it} - R_{ft}$). The 96 monthly excess returns per portfolio, over the period 2004-2011, are calculated as the equally weighted holding period returns of the individual stocks in the portfolio – selected according to the ESG ratings provided by KLD/MSCI ESG STATS – *minus* the risk-free rate for that month. The independent variables in the regressions are the MKT, SMB, HML, and MOM factors provided by the Kenneth R. French Data Library (2012).

The conjecture of this thesis is that abnormal returns will be observed due to 'shunned stock' and 'errors-in-expectations' effects, thereby inferring that the intercepts of the regressions have to be statistically different from zero; indicating either under- or outperformance, dependent on the corresponding negative or positive sign. To test this conjecture, the main statistical test used in this portfolio comparison will check whether the alpha value is different from zero. The *null* hypothesis (H_0) states that the intercept is equal to zero. This hypothesis will be tested against the *alternative* hypothesis (H_1), claiming that alpha is unequal to zero; either smaller or larger. The *null* hypothesis is tested for rejection, resulting in two possible statistical conclusions: (1) H_0 is rejected and H_1 is true, indicating abnormal returns; or (2) H_0 is not rejected, which is the less strong result not concluding anything about H_0, but only stating "that the data do not indicate that the opposite (H_1) is true" (Nieuwenhuis, 2009; p. 444).

$$H_0\colon \beta_i = 0 \qquad vs. \qquad H_1\colon \beta_i \neq 0$$

$(for\ i = 0, 1, 2, \dots, k);\ the\ test\ statistic\ is$

$$t = \frac{b_i - \beta_i}{s_{b_i}} \tag{6}$$

$which\ is\ Student\ t\ distributed\ with\ \nu = n - k - 1\ degrees\ of\ freedom.$

(Keller & Warrack, 2003; p. 667)

The regression output also contains the result of a t-test on each of the coefficients – including the intercept – accompanied by the corresponding *p*-values[11]. These two statistics summarise the outcomes of the statistical test given by formula 6. Based on these t-stats the validity of the factor loadings can be examined and, more importantly, conclusions can be drawn on the alpha values.

Before drawing conclusions on the regression output, the standard errors have been checked for the homoscedasticity and independence assumptions. Especially, the latter assumption is easily broken in time series data. The error terms are analysed by constructing scatter plots and visually looking for patterns of positive first-order autocorrelation[12]. The scatter plots for each model looked fickle and no obvious pattern has been detected. Therefore, the regressed standard errors are used in the tests and, the sometimes employed, Newey-West method is not used to recalculate standard errors (Nieuwenhuis, 2009).

To determine whether the test results are statistically significant, a confidence level has to be chosen. For all the results, the classical 0.01, 0.05 and 0.10 confidence levels are indicated. Although these significance levels are often employed in studies, the choice of these levels is quite arbitrary. Therefore, the *p*-values from the regression output are also listed. In doing so, the validity of the coefficients can be interpreted by the reader. The *p*-value states the probability that the obtained regression value departs from zero, while in reality it is equal to zero. Therefore *p*-values give the likelihood of committing a type-I error, rejecting the *null* hypothesis while it is in fact true. The overview of statistical conclusions and wrong inferences can be seen in table 5a (Bodie et al., 2009; Statman & Glushkov, 2009).

Table 5a. Statistical conclusions

	Statistical conclusion	
Actual situation	Do not reject H_0	Reject H_0
H_0 is true	Correct conclusion	Incorrect; type I
H_1 is true	Incorrect; type II	Correct conclusion

Source: Nieuwenhuis (2009)

Another measure incorporated in the analysis is the "coefficient of determination adjusted for degrees of freedom", also known as the adjusted R^2 (Keller & Warrack, 2003; p. 662). The R^2 measure is important for financial studies, since it explains which part of the risk is due to the market factors included in the model – i.e. non-diversifiable market risk. The fraction of risk that is

[11] Please refer to table III in the appendix for a sample of the unedited regression output on the portfolio of shunned stocks; all regression outputs are summarised in the tables of section 5.2.
[12] Please refer to figure IV in the appendix for an example of the scatter plots used for visually checking the independence assumption.

not included in the R^2 can be attributed to individual stocks – i.e. diversifiable firm-specific risk (Nieuwenhuis, 2009). The *adjusted* R^2, which is also presented in the regression output, is corrected for the number of independent variables and the size of the sample used. Not adjusting the R^2 for these factors might lead to an unrealistically high explanatory power of the model. Given the fact that the portfolio used are well-diversified, the smallest portfolio still holds 95 different stocks, high adjusted R^2 values are expected (Keller & Warrack, 2003).

Lastly, in general the tables with the regression results summarise the three portfolio strategies. They show the outputs for the long-only positions in both the high- and low-rated portfolios (e.g. respectively accepted and shunned stocks) and the results of a long-short strategy. This long-short strategy is added since the KLD/MSCI rated stocks are a subset of the larger CRSP universe, which is used as the benchmark in this study. A four-factor regression shows that the KLD/MSCI stocks display an alpha value of 0.89% when regressed against this broader benchmark. Thus, the KLD/MSCI subset outperforms the CRSP universe over the entire 2004-2011 period. This might result in large intercepts for the regressions on both long-only portfolios, only because the KLD/MSCI subset that is analysed in this thesis already outpaces the benchmark. The long-short position adjusts for this fact. The combined position is hedging out most exposures to the risk factors, nearly establishing a market-neutral position with absolute returns. It shows a purer comparison of the tested assumption: attaining a "pure play" on the perceived source of abnormal return and sifting out any peripheral risk factors (Bodie et al., 2009; p. 273).

5.2 The Performance of Responsible Portfolios

'Shunned stock' effect

The first hypothesis that is tested investigates the 'shunned stock' effect in which sin stocks outperform comparable stocks. For this assumption, all stocks involved in at least one controversial business area are gathered in the 'shunned' portfolio and all other KLD/MSCI rated stocks are placed in the 'accepted' portfolio.

$$Hypothesis\ 1 \qquad H_0: \alpha_{Shunned} = 0 \qquad vs. \qquad H_1: \alpha_{Shunned} \neq 0$$

The results of the regression on the first hypothesis are shown in table 5b. When looking at the adjusted R^2, the returns are explained for more than 90 percent by the risk factors in the model (also substantiated with significant F-scores). Moreover, by extending the model with the SMB, HML, and MOM factors, the adjusted R^2 increases; accrediting the higher explanatory power of multi-factor models. When looking at the intercepts for the 'accepted' portfolio, in none of the

models a significant abnormal return is observed. In this case, the absence of alpha is not that surprising. The 'accepted' portfolio contains on average more than 2000 stocks, therefore being a broad representation of the wider CRSP universe. The fact that it mimics the market benchmark would make it hard to achieve outperformance.

More important are the results for the 'shunned' portfolio, which yearly comprises on average 167 stocks. A similar trend in adjusted R^2 is observed, therefore all three benchmark models are taken into consideration. For both the CAPM and the four-factor model a significant positive intercept is reported. The single-index Jensen's alpha shows an outperformance of 4.06 percent with a *p*-value of 0.10. The four-factor alpha is lower with 3.36 percent, but the confidence level is higher with a *p*-value of 0.05. These positive abnormal returns are in line with the hypothesis that values-constraint investors forego profitable investment opportunities by refraining from perceived sinful stocks. Yet, for a purer proxy of returns for a values-driven investor, the intercepts of the long-short strategy should be examined. These imitate the returns of money invested in 'accepted' stocks and that relinquishes 'shunned' stocks. The intercepts of the long-short strategy are not significant. However, for the three- and four-factor models the underperformance according the 'shunned' stock effect is near to significant with *p*-values of 0.11.

Table 5b. Summary of the regression output for '*accepted-minus-shunned*'. Two long-only portfolios of respectively 'accepted' and 'shunned' stocks, and a third long-short portfolio strategy of buying 'accepted' stocks and selling 'shunned' stocks are shown. All portfolios are equally weighted. For the CAPM, Fama-French three-factor model and Carhart four-factor model the annualized abnormal return (intercept), factor loadings, and the adjusted R Squares over the period 2004-2011 are tabulated. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level (*p*-values within parentheses).

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Accepted							
	CAPM	1.85%	1.31***				0.91
		(0.44)	(0.00)				
	3-Factor	0.53%	1.09***	0.74***	0.25***		0.98
		(0.67)	(0.00)	(0.00)	(0.00)		
	4-Factor	0.73%	1.03***	0.79***	0.15***	-0.18***	0.99
		(0.27)	(0.00)	(0.00)	(0.00)	(0.00)	
Shunned							
	CAPM	4.06%*	1.23***				0.90
		(0.10)	(0.00)				
	3-Factor	3.16%	1.10***	0.54***	0.03		0.93
		(0.12)	(0.00)	(0.00)	(0.67)		
	4-Factor	3.36%**	1.04***	0.59***	-0.07	-0.19***	0.95
		(0.05)	(0.00)	(0.00)	(0.23)	(0.00)	
Accepted minus shunned (long-short)	CAPM	-2.21%	0.08**				0.05
		(0.23)	(0.01)				
	3-Factor	-2.63	-0.01	0.20***	0.22***		0.23
		(0.11)	(0.82)	(0.00)	(0.00)		
	4-Factor	-2.64%	-0.01	0.19***	0.22***	0.01	0.22
		(0.11)	(0.87)	(0.00)	(0.00)	(0.81)	

Data: KLD/MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007), Statman & Glushkov (2009)

The most acclaimed study on the returns of sin stocks, by Hong and Kacperczyk (2009), did find significant abnormal returns in a long-short strategy. Compared to the results in table 5b, they used a tighter definition of what comprises a sin stock. Their definition of 'sin' was only including involvement with alcohol, gambling and tobacco; which they dubbed the 'triumvirate of sin'. Their rationale behind choosing a stricter definition of sin is very plausible, since the average U.S. investor might not at all regard the weapons and military industry as controversial. Since the data of this thesis only incorporate the U.S. listed Russell3000 stocks, a 'shunned stock' effect might be stronger when only focusing on the business areas that are really perceived as controversial by U.S. investors (Chegut et al., 2011). Therefore, the portfolios are reconfigured to only incorporate alcohol-, gambling- and tobacco-related stocks in the banned portfolio. Table 5c gives the results for the reconstructed long-short strategy.

The results show that, in contrast to what was expected, the statistical power did not increase by confining the definition of sin stocks. Although the intercepts, decreasing by approximately one percentage point, confirm a more severe 'shunned stock' effect, the *p*-values have become far from significant.

Table 5c. Summary of the regression output for '*accepted-minus-shunned*' based on Hong & Kacperczyk's (2009) 'triumvirate of sin' (involvement with alcohol, gambling or tobacco). The long-short portfolio strategy buys 'accepted' stocks and sells 'shunned' stocks. The portfolio is equally weighted. For the CAPM, Fama-French three-factor model and Carhart four-factor model the annualized abnormal return (intercept), factor loadings, and the adjusted R Squares over the period 2004-2011 are tabulated. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level (*p*-values within parentheses).

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Accepted minus shunned ('triumvirate of sin')							
(long-short)	CAPM	-3.33%	-0.16**				0.05
		(0.35)	(0.01)				
	3-Factor	-3.36%	-0.14*	0.07	-0.16		0.05
		(0.34)	(0.06)	(0.65)	(0.20)		
	4-Factor	-3.70%	-0.04	-0.02	0.01	0.31***	0.30
		(0.22)	(0.52)	(0.89)	(0.91)	(0.00)	

Data: KLD/MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007), Statman & Glushkov (2009)

'Errors-in-expectations' effect

The second hypothesis sets out to find if there might be an 'errors-in-expectations' effect, in which the market not fully incorporates ESG information in stock prices. If this short-term mispricing exists, it might lead to abnormal returns when the financial benefits from good ESG performance eventually materialise. Profit-seeking investors, hoping to exploit under investigated, value-relevant factors in a quest to generate alpha, would employ a positive screening trading

strategy based on this logic. For the analysis, the generally phrased second hypothesis in section 4.2 is broken down into several individual hypotheses. Since previous research has indicated that the influence can vary heavily per parameter, the seven qualitative ESG indicators will each be examined one by one[13].

$$Hypothesis\ 2a \qquad H_0: \alpha_{COM.} = 0 \qquad vs. \qquad H_1: \alpha_{COM.} \neq 0$$

The high adjusted R^2 values are also present in the regressions on the qualitative criteria. The portfolios used in the analysis are based on the top and bottom 33 percent of stocks within the given indicator and rebalanced each year according to the up-to-date scores. The community indicator, which showed strong performance in previous studies (Kempf & Osthoff, 2007; Statman & Glushkov, 2009), is tested first. The results of the regressions can be found in table 5d.

Table 5d. Summary of the regression output for '*community*'. The two long-only portfolios of respectively high- and low-rated stocks are not displayed, since the intercepts are insignificant. The long-short portfolio strategy of buying high-rated stocks and selling low-rated stocks is shown. The portfolio is equally weighted with cut-off points of 33 percent. For the CAPM, Fama-French three-factor model and Carhart four-factor model the annualized abnormal return (intercept), factor loadings, and the adjusted R Squares over the period 2004-2011 are tabulated. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level (*p*-values within parentheses).

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Community							
Top- minus bottom-rated (long-short)	CAPM	0.12% (0.93)	-0.08*** (0.00)				0.11
	3-Factor	0.15% (0.90)	-0.10*** (0.00)	-0.07 (0.15)	0.18*** (0.00)		0.25
	4-Factor	0.14% (0.91)	-0.10*** (0.00)	-0.07 (0.14)	0.19*** (0.00)	0.01 (0.56)	0.24

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Only the output of the long-short strategy is shown, since both the high- and low-rated long-only portfolios did not deliver statistically significant intercepts. As shown in table 5d, a slight outperformance based on the community factor is witnessed – consistent with the previous studies. However, for all the benchmark models the *p*-values are anything but significant. The factor loadings for MKT and HML are significant. The former has a negative value, indicating that the low-rated community stocks are riskier compared to the top-rated. The positive HML factor loading shows a bias amongst the high-rated stocks towards value companies with a high book-to-market ratio, as compared to the low-rated portfolio.

[13] Please refer to the header in table 4c for the abbreviations of the seven qualitative indicators.

$Hypothesis\ 2b$ $H_0: \alpha_{CORP.GOV.} = 0$ $vs.$ $H_1: \alpha_{CORP.GOV.} \neq 0$

The next hypothesis is directed at the 'corporate governance' indicator (table 5e), which is often not included in performance studies due to the changing definitions over time. The four-factor regression on the low-rated corporate governance portfolio revealed an outperformance of 1.98 percent with a *p*-value of 0.05. However, this is the only significant result, which is too weak to conclude that screening portfolios solely on corporate governance strengths and concerns will deliver abnormal returns. Moreover, the negative signs of the long-short intercepts indicate that any abnormal returns would not benefit profit-seeking investors. Rather, the assumed underperformance would more likely hurt them. Finally, most factor loadings are significant. On average and compared to the companies with bad corporate governance practices, the companies with good ratings are characterised as small cap, value companies with less market exposure and high momentum.

Table 5e. Summary of the regression output for '*corporate governance'*. See table 5d for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Corporate governance							
Top- minus bottom-rated (long-short)	CAPM	-1.67% (0.34)	-0.12*** (0.00)				0.14
	3-Factor	-1.97% (0.25)	-0.18*** (0.00)	0.16** (0.02)	0.09 (0.15)		0.19
	4-Factor	-2.05% (0.21)	-0.15*** (0.00)	0.14** (0.03)	0.13** (0.03)	0.08*** (0.01)	0.24

Table 5f. Summary of the regression output for '*diversity'*. See table 5d for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Diversity							
Top- minus bottom-rated (long-short)	CAPM	-1.47% (0.37)	-0.14*** (0.00)				0.20
	3-Factor	-1.21% (0.42)	-0.13*** (0.00)	-0.21*** (0.00)	0.17*** (0.00)		0.34
	4-Factor	-1.18% (0.43)	-0.13*** (0.00)	-0.20*** (0.00)	0.15*** (0.01)	-0.02 (0.35)	0.34

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

$Hypothesis\ 2c$ $H_0: \alpha_{DIV.} = 0$ $vs.$ $H_1: \alpha_{DIV.} \neq 0$

The third qualitative ESG strength and concern indicator is 'diversity', presented in table 5f. Again, the alpha values for both long-only portfolios are not statistically different from zero. Comparable to what was observed at the 'corporate governance' indicator, the combined strategy

of shorting companies with low diversity ratings and holding the high-performers hints towards underperformance – granting that it is not statistically robust. Except for the momentum factor in the long-short strategy all factor loadings are significant.

Up to this point, the regressions on the qualitative indicators have shown very inconclusive results. This includes the 'community' indicator, which has proven to deliver significant positive alpha in earlier studies that covered another time span and the smaller S&P500 and DSI400 universes. Following the results of these previous studies, other qualitative indicators that should show outperformance are 'employee relations' and to some extent the 'human rights' and 'environment' factors. These three, accompanied by the 'product' indicator will be examined in more detail below (Kempf & Osthoff, 2007; Statman & Glushkov, 2009).

$Hypothesis\ 2d$ $\quad H_0: \alpha_{EM.REL.} = 0 \quad vs. \quad H_1: \alpha_{EM.REL.} \neq 0$

The academic evidence on the positive effects of social screening have up till now always been strongest for the 'community' and the 'employee relations' indicators. In contrast to the earlier examined community factor, the regression results on employee relations (table 5g) show at least one postive intercept. But again, the long-short strategy that mimics the returns of a profit-seeking investor does not provide a source for abnormal returns. Combining the negative – insignificant – intercepts of the long-short strategy with the slightly significant four-factor alpha of 1.57 percent for the bottom-rated portfolio, it can be judged that if any effect at all is observed, it would imply underperformance when overinvesting in companies with good employee relations.

$Hypothesis\ 2e$ $\quad H_0: \alpha_{ENV.} = 0 \quad vs. \quad H_1: \alpha_{ENV.} \neq 0$

Table 5h displays the regression results for the 'environmental' indicator. When looking at the portfolio of bottom-rated stocks, positive alphas are found. Based on the three- and four-factor intercepts, an outperformance of around 2.50 percent is observed, with the declining *p*-values indicating statistically stronger results. Unfortunately for profit-seeking investors, the intercept values for the long-only portfolio of top-rated environmental stocks are far from significant. Despite, or perhaps because of, the fact that environmental sustainability is a really tangible and measureable indicator, no positive abnormal returns can be achieved by only investing in companies with good environmental performance. Moreover, the displayed alpha values for the top-portfolio are even smaller than the portfolio of low-scorers on environment. The long-short strategy does not yield significant intercepts, but unsurprisingly it shows a tendency of underperformance for environment-minded investors; *p*-values range from 0.14 to 0.16.

Table 5g. Summary of the regression output for '*employee relations'*. Two long-only portfolios of respectively high- and low-rated stocks, and a third long-short portfolio strategy of buying high-rated stocks and selling low-rated stocks are shown. All portfolios are equally weighted with cut-off points of 33 percent. For the CAPM, Fama-French three-factor model and Carhart four-factor model the annualized abnormal return (intercept), factor loadings, and the adjusted R Squares over the period 2004-2011 are tabulated. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level (*p*-values within parentheses).

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Employee							
relations	CAPM	1.33%	1.23***				0.91
top-rated		(0.56)	(0.00)				
	3-Factor	0.17%	1.03***	0.62***	0.31***		0.97
		(0.90)	(0.00)	(0.00)	(0.00)		
	4-Factor	0.34%	0.98***	0.67***	0.22***	-0.16***	0.99
		(0.69)	(0.00)	(0.00)	(0.00)	(0.00)	
Employee							
relations	CAPM	2.74%	1.36***				0.90
bottom-rated		(0.29)	(0.00)				
	3-Factor	1.35%	1.14***	0.80***	0.18***		0.97
		(0.37)	(0.00)	(0.00)	(0.00)		
	4-Factor	1.57%*	1.07***	0.85***	0.07**	-0.20***	0.99
		(0.08)	(0.00)	(0.00)	(0.04)	(0.00)	
Top- minus bottom-rated							
(long-short)	CAPM	-1.41%	-0.12***				0.25
		(0.25)	(0.00)				
	3-Factor	-1.19%	-0.11***	-0.18***	0.13***		0.40
		(0.28)	(0.00)	(0.00)	(0.00)		
	4-Factor	-1.23%	-0.09***	-0.19***	0.15***	0.02**	0.42
		(0.26)	(0.00)	(0.00)	(0.00)	(0.04)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Table 5h. Summary of the regression output for '*environment'*. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Environment							
top-rated	CAPM	1.56%	1.23***				0.91
		(0.50)	(0.00)				
	3-Factor	0.39%	1.03***	0.63***	0.28***		0.97
		(0.78)	(0.00)	(0.00)	(0.00)		
	4-Factor	0.59%	0.98***	0.68***	0.18***	-0.18***	0.99
		(0.53)	(0.00)	(0.00)	(0.00)	(0.00)	
Environment							
bottom-rated	CAPM	3.59%	1.36***				0.92
		(0.14)	(0.00)				
	3-Factor	2.31%*	1.15***	0.73***	0.19***		0.97
		(0.09)	(0.00)	(0.00)	(0.00)		
	4-Factor	2.50%***	1.10***	0.77***	0.10***	-0.17***	0.99
		(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	
Top- minus bottom-rated							
(long-short)	CAPM	-2.03%	-0.12***				0.22
		(0.14)	(0.00)				
	3-Factor	-1.92%	-0.12***	-0.10*	0.09*		0.25
		(0.15)	(0.00)	(0.07)	(0.06)		
	4-Factor	-1.91%	-0.12***	-0.10*	0.09*	0.00	0.24
		(0.16)	(0.00)	(0.08)	(0.08)	(0.91)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

When analysing the factor loadings of the regression on the long-short portfolio returns, it shows a tilt for the bottom-rated stock portfolio towards large cap, value companies (significant at the 0.10 level). This might be retraceable to the characteristics of the companies in the 'old' industries, like energy exploration and mining, which are often heavy polluters, but well performing companies.

$Hypothesis\ 2f$ $H_0: \alpha_{HUM.RIG.} = 0$ $vs.$ $H_1: \alpha_{HUM.RIG.} \neq 0$

The sixth hypothesis on positive screening covers the 'human rights' indicator (table 5i). Again, the only strategy that shows significant results is the long-only bottom-rated. The three- and four-factor intercepts are 2.38 and 2.57 percent. The former is significant at the 0.10 level and the latter at the 0.01 level. No reliable conclusions can be made on the alpha values of the top-rated portfolio. Therefore, the long-short portfolio is heavily influenced by the outperformance of the low-rated human rights portfolio. The alpha values for the combined strategy are negative with *p*-values varying around 0.14. No inferences can be made on this basis, but it fits in the general pattern in which the bottom-rated stocks tend to outperform the top-rated stocks.

$Hypothesis\ 2g$ $H_0: \alpha_{PRO.} = 0$ $vs.$ $H_1: \alpha_{PRO.} \neq 0$

The regression results for 'product', the last qualitative indicator that is studied separately, are found in table 5j. The abovementioned trend in the alpha values is also witnessed amongst the stocks listed according to the product indicator. The bottom-rated portfolio shows significant outperformance, which cascades down into the alpha values for the long-short strategy. Yet, with regards to product, the intercept values are very significant. All three benchmark models indicate outperformance for the low-rated portfolio, with annual abnormal returns of 3.75, 2.72, and 2.92 percent for respectively the CAPM, Fama-French, and Carhart alphas. The corresponding *p*-values are 0.08, 0.05, and 0.00; indicating strong empirical results. The long-short strategy also has robust intercept, showing underperformance in excess of three percent; the matching alpha values are significant at the 0.05 level. This indicates a very unbeneficial effect for a profit-seeking responsible investor, since the results in table 5j show that this strategy would deliver negative absolute returns over the period 2004-2011.

Table 5i. Summary of the regression output for '*human rights'*. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Human rights top-rated	CAPM	1.45%	1.23***				0.91
		(0.53)	(0.00)				
	3-Factor	0.27%	1.03***	0.64***	0.29***		0.97
		(0.85)	(0.00)	(0.00)	(0.00)		
	4-Factor	0.45%	0.97***	0.68***	0.19***	-0.17***	0.98
		(0.63)	(0.00)	(0.00)	(0.00)	(0.00)	
Human rights bottom-rated	CAPM	3.64%	1.36***				0.92
		(0.13)	(0.00)				
	3-Factor	2.38%*	1.16***	0.72***	0.18***		0.97
		(0.09)	(0.00)	(0.00)	(0.00)		
	4-Factor	2.57%***	1.10***	0.77***	0.08**	-0.17***	0.99
		(0.01)	(0.00)	(0.00)	(0.02)	(0.00)	
Top- minus bottom-rated (long-short)	CAPM	-2.19%	-0.13***				0.20
		(0.13)	(0.00)				
	3-Factor	-2.11%	-0.13***	-0.08	0.11**		0.24
		(0.14)	(0.00)	(0.15)	(0.03)		
	4-Factor	-2.11%	-0.13***	-0.08	0.11**	0.00	0.23
		(0.14)	(0.00)	(0.15)	(0.04)	(0.90)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Table 5j. Summary of the regression output for '*product'*. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Product top-rated	CAPM	0.73%	1.25***				0.90
		(0.77)	(0.00)				
	3-Factor	-0.62%	1.02***	0.74***	0.29***		0.97
		(0.64)	(0.00)	(0.00)	(0.00)		
	4-Factor	-0.45%	0.97***	0.78***	0.21***	-0.15***	0.99
		(0.63)	(0.00)	(0.00)	(0.00)	(0.00)	
Product bottom-rated	CAPM	3.75%*	1.31***				0.93
		(0.08)	(0.00)				
	3-Factor	2.72%**	1.13***	0.57***	0.22***		0.97
		(0.05)	(0.00)	(0.00)	(0.00)		
	4-Factor	2.92%***	1.08***	0.61***	0.12***	-0.18***	0.99
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	
Top- minus bottom-rated (long-short)	CAPM	-3.02%**	-0.05**				0.04
		(0.04)	(0.03)				
	3-Factor	-3.34%**	-0.11***	0.18***	0.07		0.14
		(0.02)	(0.00)	(0.00)	(0.13)		
	4-Factor	-3.37%**	-0.10***	0.17***	0.09*	0.02	0.14
		(0.02)	(0.00)	(0.00)	(0.09)	(0.33)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Overall, the risk factors in the benchmark models managed to explain most of the variation in the portfolio returns, with adjusted R^2 values usually in excess of 0.90. Concerning the hypotheses tested, the first three – 'community' (2a), 'corporate governance' (2b), and 'diversity' (2c) – did not deliver any statistical results. Therefore, their *null* hypotheses cannot be rejected and the existence of abnormal returns for these three indicators is not proven. The other four indicators showed stronger results for the long-only bottom-rated portfolios. The 'employee relations' four-factor intercept hints towards outperformance for low-rated companies. An outcome that is even more significant for the 'environment' and 'human rights' indicators. Furthermore, the outperformance of the bottom-rated portfolio is also observable in the negative alpha values of the long-short strategy; alpha values which approached statistical significance at the 0.10 level. However, the results are not strong enough to reject the *null* hypothesis. Therefore, also for hypotheses 2d, 2e, and 2f the alpha values are not distinguishable from zero and it has to be concluded that no abnormal returns are found.

Only the evidence on the 'product' indicator is significant enough to rejected the *null* hypothesis and accept the *alternative* hypothesis with 95 percent confidence. Nevertheless, in contrast to what was assumed upfront, the observed abnormal returns are not positive. The long-short intercepts all carry negative signs, indicating underperformance. Although only hypothesis 2g provides statistically significant evidence on the underperformance for responsible investors, the similar pattern is apparent amongst all other indicators.

Another testable hypothesis with regards to the qualitative indicators remains. The overall ESG scores, as calculated by the relative method in formula 2 and 3, are also used to construct eight yearly high- and low-rated portfolios. The conjecture remains the same and is represented by hypothesis 2h. The results of the regression analysis on the return series can be found in table 5k.

$$Hypothesis\ 2h \qquad H_0: \alpha_{ESG} = 0 \qquad vs. \qquad H_1: \alpha_{ESG} \neq 0$$

The assumption is not supported by the results, at least, not the sought after 'errors-in-expectations' effect. Parallel to the individual indicators, no significant positive outperformance is reported; not for the long-only top-rated portfolio and certainly not for the long-short strategy. However, again the opposite is true; the bottom-rated portfolio displays a positive four-factor abnormal return of 2.27 percent and a *p*-value of 0.01, reaching a confidence level of almost 99 percent. When shifting the attention to the long-short strategy, statistically strong intercept values are found. But in line with the outperformance for the poorly rated ESG performers, also this long-short strategy will result in underperformance for the profit-seeking responsible investor.

Nonetheless, despite the fact that the direction of the abnormal returns is not as expected, for hypothesis 2h the *null* form can still be rejected. The alpha values for the long-short strategy indicate abnormal returns; only negative instead positive performance, but certainly statistically different from zero. The CAPM Jensen's alpha of -2.62 percent is not statistically significant with a *p*-value of 0.11. On the other hand, the three- and four-factor intercepts of correspondingly -2.75 and -2.78 percent have *p*-values of 0.08; making them significant for at least the 0.10 level. Moreover, these benchmark models are characterized by a higher adjusted R^2. As a result, it can be concluded with 90 percent confidence that the *alternative* hypothesis holds, indicating negative abnormal returns for profit-seeking investors using a strategy based on overall ESG scores.

Table 5k. Summary of the regression output for '*overall ESG score*'. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Top-rated							
overall	CAPM	0.64%	1.23***				0.88
		(0.81)	(0.00)				
	3-Factor	-0.70%	1.00***	0.72***	0.33***		0.96
		(0.65)	(0.00)	(0.00)	(0.00)		
	4-Factor	-0.51%	0.94***	0.77***	0.23***	-0.17***	0.98
		(0.65)	(0.00)	(0.00)	(0.00)	(0.00)	
Bottom-rated							
overall	CAPM	3.26%	1.36***				0.92
		(0.17)	(0.00)				
	3-Factor	2.05%	1.16***	0.69***	0.18***		0.97
		(0.17)	(0.00)	(0.00)	(0.00)		
	4-Factor	2.27%**	1.10***	0.74***	0.07**	-0.20***	0.99
		(0.01)	(0.00)	(0.00)	(0.03)	(0.00)	
Top-rated minus bottom-rated							
(long-short)	CAPM	-2.62%	-0.13***				0.17
		(0.11)	(0.00)				
	3-Factor	-2.75%*	-0.16***	0.03	0.15***		0.21
		(0.08)	(0.00)	(0.59)	(0.01)		
	4-Factor	-2.78%*	-0.15***	0.03	0.16***	0.02	0.21
		(0.08)	(0.00)	(0.66)	(0.00)	(0.38)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

'Best-in-class' screening

The third and last overall hypothesis stated in section 4.2 employs an industry-balanced portfolio, also referred to as a best-in-class technique. For this hypothesis, the overall ESG scores are used again, but now the stocks have been arranged in ten separate industries. In doing so, a diversified sector allocation is guaranteed and no single – clean or dirty – industry is overrepresented in the sample. The best-in-class (*BIC*) screening strategy yielded the highest positive abnormal returns in previous studies. Given the results from the foregoing hypotheses it is questionable whether regrouping the stocks according to a balanced sector allocation will suddenly

fulfill the positive outperformance conjecture. The results from the regression on hypothesis 3 are shown in table 5l and will be discussed below.

$Hypothesis\ 3 \qquad H_0: \alpha_{BIC} = 0 \qquad vs. \qquad H_1: \alpha_{BIC} \neq 0$

Offsetting the best performers, in each industry, against the worst performers, has not resulted in major changes compared to the normal overall ESG regressions tested by hypothesis 2h. The long-only top-rated portfolio has insignificant intercepts. The long-only bottom-rated portfolio shows outperformance, which becomes smaller but statistically stronger for the multi-factor models. Finally, the long-short strategy – most important for the hypothesis test – delivers negative abnormal returns for a responsible investor. The alpha values are approximately -2.80 percent with *p*-values ranging from 0.04 to 0.02. In that sense, the best-in-class screening strategy points in the same direction as the earlier tests. Taking into account the industries does however make a difference. On the basis of the observed abnormal returns, the *null* hypothesis can be rejected. This time with at least 96 percent confidence, according to the CAPM and 98 percent for the multi-factor models. This is in line with the conclusions of previous research; using the best-in-class methodology makes the results statistically more robust.

Table 5l. Summary of the regression output for '*best-in-class screening*'. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Top-rated							
overall	CAPM	0.77%	1.25***				0.90
best-in-class		(0.75)	(0.00)				
	3-Factor	-0.52%	1.03***	0.71***	0.29***		0.97
		(0.70)	(0.00)	(0.00)	(0.00)		
	4-Factor	-0.34%	0.97***	0.75***	0.19***	-0.17***	0.99
		(0.71)	(0.00)	(0.00)	(0.00)	(0.00)	
Bottom-rated							
overall	CAPM	3.48 %	1.33***				0.92
best-in-class		(0.14)	(0.00)				
	3-Factor	2.30%	1.14***	0.68***	0.14***		0.97
		(0.12)	(0.00)	(0.00)	(0.01)		
	4-Factor	2.51%***	1.08***	0.73***	0.04	-0.19***	0.99
		(0.01)	(0.00)	(0.00)	(0.27)	(0.00)	
Top-rated minus bottom-rated							
(long-short)	CAPM	-2.72%**	-0.08***				0.11
		(0.04)	(0.00)				
	3-Factor	-2.82%**	-0.11***	0.03	0.14		0.19
		(0.02)	(0.00)	(0.61)	(0.00)		
	4-Factor	-2.85%**	-0.11***	0.02	0.16***	0.02	0.19
		(0.02)	(0.00)	(0.69)	(0.00)	(0.33)	

Note: Top- and bottom 33 percent performers of each industry are selected in the respective portfolios.
Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

5.3 Subsample analysis

Briefly reviewing the results of the hypothesis testing, reveals that a 'shunned stock' effect is recognised, although the intercepts of the long-short strategy exploiting it fell just outside the conventional confidence levels. Subsequently, in line with previous studies, not all individual indicators yield abnormal returns. However, when the 'product' rating is used to construct a portfolio, a significant alpha value can be reached. The same is true for the long-short strategy using the overall ESG score – both for the normal and best-in-class approach. Yet, in constrast to prior academic work, the abnormal returns do not indicate positive outperformance. All portfolios were built on the conjecture that good ESG performance would result in positive alpha values, but the opposite has proven to be the case. The significant alpha values, given by the regression models, all indicate negative abnormal returns and thus underperformance for the responsible investor. To check whether these results change when altering the inputs, this section puts forth a number of additional tests. Note, the data for the 'shunned stock' hypothesis has already been altered, by only including alcohol-, gambling-, and tobacco-related stocks in the controversial or 'shunned' portfolio. That additional test did not deliver improved results.

Alternative cut-off points

First, a different cut-off point is applied. The 33 percent boundaries used before are restricted to 5 percent, only including the best and worst performers on a given indicator into the portfolios. In previous studies, these 5 percent cut-off points are the smallest fractions used. It is said, that the larger the ESG performance differential between the high- and low-rated portfolios, the more significant the financial results for a long-short strategy will become. Tables V, VI, and VII in the appendix show the renewed results for the 'community', 'employee relations' and 'product' indicators. The first two are retested because their alpha values are insignificant in this study, yet previous research did find statistically strong outperformance. Therefore, it is checked whether tightening the cut-off point will change the significance of the results. 'Product' is the only qualitative indicator that has a significant intercept for the long-short strategy, therefore that result is also retested with a different cut-off point.

Summarising the result of the regressions, the effect of tightening the cut-off points is mixed. First, for 'community' the regression results further deviate from the conjecture that high-rated companies outperform. The alpha value for the high-rated 'community' portfolio decreases, and its low-rated portfolio's alpha increases. As a result, the positive long-short intercept that hinted towards outperformance – the only one witnessed in the entire analysis above – has disappeared when using the 5 percent bounds. In addition, it has to be taken into consideration that the intercepts are still far from significant and therefore not the basis of solid conclusions.

Secondly, the other two indicators do show to the desired effect. The intercepts for the high-rated portfolios increase, while the low-rated alphas decline. In total, this leads to a smaller difference between the performance of both portfolios. A fact demonstrated by the long-short alpha that becomes less negative. Yet again, since the regressions with the renewed boundaries are not significant, no real inferences on the abnormal returns can be made. Changing the cut-off point does therefore not strengthen the results for the abnormal returns.

Lastly, the factor loadings do remain significant and provide some solid information. For instance, the altered cut-off points have declined the small cap bias witnessed in the high-rated employee relations portfolio. It is an unproven statement, but it could very well be that the five percent cut-off only captures the larger, better-institutionalised companies, that score high on this dimension.

Russell3000 versus S&P500 universe

This thesis has used a broader data set as compared to other studies; i.e. the Russell3000 universe versus the often researched S&P500 constituents. To see if this choice has any profound effects on the results, an additional regression only using S&P500 stocks is performed. This analysis is executed using the best-in-class methodology and the overall ESG scores. The regression results can be found in table 5m. Immediately jumping to the conclusion, the results are consistent with the previous tests. The S&P500-based portfolios show no significant intercepts for the long-only high-rated portfolio, strong – statistical – outperformance for the long-only low-rated portfolio, and significant negative alpha values for the long-short strategy. Moreover, the underperformance following the long-short strategy is stronger than observed before. The *p*-values are very low, enabling a statistical conclusion with at least 98 percent confidence. Perhaps, the fact that these results are aligned with the larger Russell3000 data set is not surprising when looking at figure VIII in the appendix, which shows the performance of both indices from 2004 to 2011. The development of both stock universes goes hand-in-hand, the performance of the indices not deviating far from each other.

Finally, the factor loadings remain significant as well. In this case it is interesting to examine the SMB factor because it is based on market capitalisation. By confining the large Russell3000 stock universe, and only looking at the S&P500 constituents, this factor should be affected. Congruent with this modification and line of reasoning, the small cap bias for both long-only portfolios is decreased substantially.

Table 5m. Summary of the subsample regression output for '*best-in-class screening S&P500 (33%)*'. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Top-rated overall best-in-class S&P500	CAPM	0.44%	1.13***				0.92
		(0.82)	(0.00)				
	3-Factor	-0.08%	1.03***	0.26***	0.20***		0.94
		(0.96)	(0.00)	(0.00)	(0.00)		
	4-Factor	0.13%	0.97***	0.32***	0.09*	-0.19***	0.97
		(0.92)	(0.00)	(0.00)	(0.05)	(0.00)	
Bottom-rated overall best-in-class S&P500	CAPM	3.79%**	1.18***				0.95
		(0.02)	(0.00)				
	3-Factor	3.44%**	1.10***	0.16***	0.20***		0.96
		(0.02)	(0.00)	(0.01)	(0.00)		
	4-Factor	3.62%***	1.05***	0.20***	0.10***	-0.17***	0.98
		(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	
Top-rated minus bottom-rated (long-short)	CAPM	-3.35**	-0.05*				0.03
		(0.02)	(0.06)				
	3-Factor	-3.52%**	-0.07**	0.10*	0.00		0.04
		(0.01)	(0.02)	(0.07)	(0.98)		
	4-Factor	-3.49%**	-0.08**	0.11*	-0.01	-0.03	0.04
		(0.02)	(0.01)	(0.06)	(0.81)	(0.31)	

Note: Top- and bottom 33 percent performers of each industry are selected in the respective portfolios.
Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Chronological subsets

The previous adaptations have not falsified the results. On the contrary, the latter subsample regression, using only S&P500 constituents, even reconfirmed the results in section 5.2. As a final check, the sample period is broken down into two subsamples of equal length, ranging from 2004-2007 and 2008-2011. Since this will halve the return series to a mere 48 observations, it is likely that the results will be less significant. Furthermore, it is not possible to compare this subsample analysis with earlier research, since the time spans do not overlap[14]. The results are shown in table 5n.

Overall, the temporal subsample results show that the first period 2004-2007 displays mainly negative values for the intercepts of both long-only strategies. The opposite, positive alpha values, is observed in the 2008-2011 period. In the first period, the top-rated long-only portfolio even reaches significant alpha values with a minimum of -5.49 percent and a 0.02 *p*-value. The bottom-rated portfolio does not have statistically strong alpha values in that time span. In the second period, these observations are different. The top-rated portfolio has insignificant alpha values and the bottom-rated long-only portfolio shows positive outperformance by at least 2.56 percent, with a 0.10 *p*-value, and at most a stunning 8.62 percent, with a *p*-value of 0.04. Yet, the

[14] Kempf & Osthoff (2007) and Statman & Glushkov (2009) only perform a subsample analysis from 1998-2004 and 2000-2007.

intercept values for the long-short strategy are not statistically strong enough to reject the *null* hypothesis for either period.

Relating these subsample results to the overall analysis, the general pattern in the abnormal returns is upheld; the top-rated portfolio performs worse than the bottom-rated portfolio. The implication for a responsible investor, investing in high-rated stocks and abstaining from low-rated stocks, is financial underperformance.

Table 5n. Summary of the subsample regression output for '*overall ESG score*'. The portfolios are tested over two separate four-year periods, spanning 2004-2007 and 2008-2011. See table 5g for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Panel A: 2004-2007							
Top-rated overall	CAPM	-5.49%**	1.25***				0.85
		(0.02)	(0.00)				
	3-Factor	-2.85%*	0.93***	0.60***	0.04		0.95
		(0.05)	(0.00)	(0.00)	(0.52)		
	4-Factor	-1.76%	0.94***	0.65***	0.06	-0.16***	0.96
		(0.18)	(0.00)	(0.00)	(0.34)	(0.00)	
Bottom-rated overall	CAPM	-2.65%	1.43***				0.89
		(0.24)	(0.00)				
	3-Factor	-0.10%	1.10***	0.64***	0.11**		0.98
		(0.92)	(0.00)	(0.00)	(0.02)		
	4-Factor	0.31%	1.10***	0.66***	0.12**	-0.06*	0.98
		(0.76)	(0.00)	(0.00)	(0.02)	(0.07)	
Top-rated minus bottom-rated (long-short)	CAPM	-2.85%*	-0.18***				0.18
		(0.09)	(0.00)				
	3-Factor	-2.74%	-0.17**	-0.05	-0.07		0.16
		(0.13)	(0.02)	(0.56)	(0.41)		
	4-Factor	-2.07%	-0.16**	-0.02	-0.06	-0.10*	0.19
		(0.25)	(0.03)	(0.85)	(0.47)	(0.09)	
Panel B: 2008-2011							
Top-rated overall	CAPM	6.68%	1.24***				0.90
		(0.16)	(0.00)				
	3-Factor	2.79%	0.99***	0.79***	0.41***		0.97
		(0.26)	(0.00)	(0.00)	(0.00)		
	4-Factor	1.23%	0.94***	0.82***	0.30***	-0.15***	0.98
		(0.51)	(0.00)	(0.00)	(0.00)	(0.00)	
Bottom-rated overall	CAPM	8.62%**	1.35***				0.93
		(0.04)	(0.00)				
	3-Factor	4.79%*	1.17***	0.71***	0.19**		0.97
		(0.10)	(0.00)	(0.00)	(0.02)		
	4-Factor	2.56%*	1.09***	0.77***	0.03	-0.22***	0.99
		(0.10)	(0.00)	(0.00)	(0.47)	(0.00)	
Top-rated minus bottom-rated (long-short)	CAPM	-1.97%	-0.12***				0.15
		(0.49)	(0.00)				
	3-Factor	-2.00%	-0.18***	0.07	0.22***		0.26
		(0.46)	(0.00)	(0.47)	(0.01)		
	4-Factor	-1.33%	-0.16***	0.06	0.26***	0.07*	0.30
		(0.62)	(0.00)	(0.56)	(0.00)	(0.08)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

6. Discussion

6.1 Conclusions

Hypothesis 1: 'Shunned stock' effect

The first hypothesis test conducted in this thesis compared a portfolio with controversial stocks against all other KLD/MSCI rated stocks. As already discussed in the results, it was not surprising that the 'accepted' portfolio, which is a broad representation of the market, did not show abnormal returns. The most important finding is the positive outperformance for the 'shunned' portfolio, which shows significant alpha values for the CAPM and Carhart four-factor models. Relating this back to the effect of shunning sin stocks, which is discussed in the theoretical review, it implies that the idiosyncratic risk does indeed play a significant role for portfolios of sin stocks.

The results are in line with the existing academic evidence, as previous studies have already extensively tested this 'shunned stock' effect and proven that it exists. By employing negative screens, the values-driven investor is only allowed to buy 'accepted' stocks, which are congruent with his or her beliefs. Although the 'accepted' stocks do not out- or underperform the market, the values-driven investor foregoes the positive excess risk-adjusted returns exhibited by the 'shunned' stocks. This results in a suboptimal return for the values-driven investor, which is shown by the long-short strategy. The *p*-values (0.11) for the long-short alphas in this study are just outside the classical confidence levels; therefore, the *null* hypothesis has not been rejected. Nevertheless, combining the findings and the significant results from previous research it can be assumed that values-driven responsible investors are worse off by incorporating their non-financial beliefs into stock selection (Hong & Kacperczyk, 2009; Statman & Glushkov, 2009).

Hypotheses 2 and 3: 'Errors-in-expectations' effect and best-in-class screening

To test whether profit-seeking responsible investors can benefit from investing in top-rated ESG companies, several hypotheses were tested. The KLD/MSCI data provided seven ESG indicators, which all have been tested individually. Eventually, only for the 'product' indicator the *null* hypothesis is rejected. A long-short strategy based on the 'product' indicator delivers negative abnormal returns to a profit-seeking investor, evidenced by the intercepts of all three benchmark models. This outcome can be attributed to the outperformance of the low-rated long-only portfolio in comparison to the high-rated long-only portfolio. Even though the alpha values for the other six qualitative indicators are statistically indistinguishable from zero, the results also show a tendency towards underperformance.

In addition, the overall ESG scores are also studied. Two methods for portfolio formation were used; firstly, similar to the individual indicators, by incorporating the companies constituting

the top- and bottom 33 percent of the overall ESG scores (i.e. hypothesis 2). Secondly, by employing the best-in-class methodology that preserves a balanced representation of industries within the portfolio (i.e. hypothesis 3). For both approaches, the results are in line with the conclusion for the individual indicators. The long-only bottom-rated portfolios show significant, or near to significant, outperformance, which accordingly also results in statistically strong abnormal returns for the long-short strategies. By employing the best-in-class methodology, the negative abnormal returns for the long-short strategy even become more robust.

Though spanning another time period, earlier studies (Kempf & Osthoff, 2007; Statman & Glushkov, 2009) did manage to find an 'errors-in-expectations' effect in which superior ESG performance was linked to financial outperformance. However, following the results of this thesis, the conclusion should be that profit-seeking investors are actually *not* served by using positive screens. For all tests conducted in this study – including the additional analyses with different cut-off points, only examining the S&P500 constituents and the temporal subsamples – the results show that overinvesting in companies that perform high on ESG criteria will lead to inferior financial performance. A positive link between corporate social performance and corporate financial performance is not found, therefore the conclusion deviates from the results found in previous research employing similar methodologies[15].

The reason behind this opposing evidence is not clear. Derwall et al. (2011), who coined the 'errors-in-expectations' hypothesis, already made inferences about the longevity of abnormal returns for responsible investors. They believed that the underperformance according to the 'shunned stock' effect should persist, but they also already stated that the outperformance due to underestimation of ESG benefits should be more short-lived. Based on the results of this thesis and the fact that responsible investment has become a mainstream investment practice, it might very well be that market has corrected the errors in expectations regarding good ESG performance, which were witnessed in earlier studies. With the demand for stocks with a good ESG record rising, and inferring from the significant underperformance evidenced in this thesis, it can be assumed that the market actually overcorrected any previously underestimated prices.

Having discussed all three main hypotheses, the effects of different ESG-based screening strategies on achieving excess risk-adjusted returns have become apparent. To finalise the conclusion for this thesis, the answer for the overall research question can be specified. Building on the results, negative, positive, and best-in-class screening strategies deliver negative risk-

[15] Given the contradicting nature of the test results compared to previous research, a supplementary regression comparing the returns of the MSCI KLD 400 Social index and the S&P500 index has been performed. A long-short strategy of the MSCI KLD 400 minus the S&P500 did however not result in significant alpha values. The intercepts were slightly negative though, showing underperformance of socially screened stocks over a conventional, unscreened stock universe; in line with these conclusions (MSCI; Yahoo! Finance, 2012).

adjusted returns. Therefore, it has become clear that screening strategies have an influence on abnormal returns, but the consequential underperformance is probably not the effect a responsible investor desires. The values-driven investors, constituting a niche market within responsible investment, look to satisfy self-transcendent values. Therefore, for this subgroup the underperformance according to the 'shunned stock' effect might be compensated by non-pecuniary benefits. However, the mainstream profit-seeking responsible investor does not display value-expressive behaviour; for them, the underperformance resulting from positive screening is not generating the desired positive abnormal returns and results in doing financially not well.

6.2 Limitations

Although the analysis in this thesis has been carefully conducted, not all contingencies could be considered. First of all the data analysis is constrained by the availability of ESG rating information. As discussed before, the KLD/MSCI ESG STATS was the only freely available data source. Perhaps the fact that these ratings are at the disposal of a random master student already signals that they are not unique enough to offer the extra-financial information needed to achieve positive abnormal returns. Therefore, the results of this thesis cannot speak for all other proprietary ESG rating methodologies used in investment research and are only valid for the MSCI ratings on the U.S. listed companies. Secondly, the ESG ratings on the Russell3000 companies only ranged from 2003 up until now, which enabled a financial analysis from 2004 to 2011. A longer time horizon and correspondingly more monthly return observations would make the outcomes of the analysis more robust. Moreover, weekly return data and momentum factors were not immediately accessible. Incorporating weekly returns in the analysis might also contribute to the significance of the analysis.

To limit the breadth of the portfolio analysis, only an equally weighted scheme was used to calculate the monthly portfolio returns. Some of the previous studies also applied a value-weighted method. Although their results encouraged the usage of equally weighted returns, the application of both techniques could enhance the results. Especially when noting that the conclusions in this thesis deviate from the prior work. Moreover, for the Russell3000 universe an equally weighted scheme might put too much emphasize on the returns of small cap stocks. Furthermore, the portfolio returns have not been corrected for management fees or transaction costs. When a positive abnormal return is displayed, adding these checks will give insight on whether the returns remain positive after accounting for costs.

Finally, as already touched upon in the literature review, accounting-based measures of financial returns tend to show a stronger relationship with the performance on ESG factors. Although accounting-based measures inherently are backward looking and not able to capture the

long-term character of ESG investments, they still might aid in explaining the negative abnormal returns witnessed in this study. Based on the results, no conclusion can be reached on the source of the negative abnormal returns. The analyses in this thesis are unable to state whether superior ESG performance simply does not deliver positive cash flows or if the underperformance is due to a market mispricing. Additionally, market-based performance measures are susceptible to many uncontrollable factors that provoke noise in the results. Using accounting-based information, such as the impact on cash flows or by incorporating earnings announcements, might be able to establish why portfolios of high-rated stocks have underperformed over the 2004-2011 period.

6.3 Recommendations

Since the conclusions on the 'shunned stock' effect are in line with academic consensus, the recommendations mainly concern the 'errors-in-expectations' effect – the possible financial outperformance based on integrating ESG indicators. The previous research and this thesis have mainly looked into the relationship between ESG and stock price performance. Despite the mixed evidence, it can be assumed that extra-financial information and motivations have an effect on abnormal returns. However, the source of such abnormal returns has not been widely studied. Mănescu (2011) was one of the few scholars focussing on the existence of a (non-)sustainability risk factor. Additional research can substantiate the conclusion that such a risk factor or residual variance does not exist and abnormal returns are due to mispricing.

Furthermore, previous studies often reached inconclusive results. With the current segmentation in the previously homogeneous group of responsible investors and the multiple effects in play, it is worth investigating whether in the foregoing studies the 'shunned stock' and 'errors-in-expectations' effects may have cancelled each other out; causing insignificant evidence. This thesis could not substantiate such a claim, since a positive 'errors-in-expectations' effect was not found, but a meta-analysis on this 'net no effect' assumption could add value to the understanding of the relationship between stock performance and screening intensity.

Finally, event studies investigating the release of new or revised ESG information on the market would be another interesting avenue for research. It would explain whether the market immediately incorporates negative ESG information and underestimates the impact of positive ESG indicators. Commercial ESG rating agencies, for instance 'Sustainalytics', publish quarterly updates on a company's ESG performance. Based on the changed ESG scores in these updates, the impact on stock prices can be investigated. Unfortunately, such information is based on proprietary rating methodologies, which is often not freely available.

References

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

Becchetti, L. & Ciciretti, R. (2009) Corporate social responsibility and stock market performance. *Applied Financial Economics, 19,* 1283-1293.

Bello, Z.Y. (2005) Socially responsible investing and portfolio diversification. *The Journal of Financial Research, 28,* 41-57.

Bengtsson, E. (2008) A history of Scandinavian socially responsible investing. *Journal of Business Ethics, 82,* 969-983.

Blowfield, M. & Murray, A. (2008) *Corporate responsibility: A critical introduction.* New York, NY: Oxford University Press.

Bodie, Z., Kane, A. & Marcus, A.J. (2009) *Investments* (8th ed.). New York, NY: McGraw-Hill.

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

Campbell, J.Y., Lo, A.W. & MacKinlay, A.C. (1997) *The econometrics of financial markets.* Princeton, NJ: Princeton University Press.

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

Carroll, A.B. (1999) Corporate social responsibility: Evolution of a definitional construct. *Business & Society, 38,* 268-295.

Chegut, A., Schenk, H. & Scholtens, B. (2011) Assessing SRI fund performance research: Best practices in empirical analysis. *Sustainable Development, 19,* 77-94.

CRSP (2012) *The Center for Research in Security Prices.* Retrieved from http://wrds-web.wharton.upenn.edu/wrds/

References

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

Diltz, J.D. (1995a) Does social screening affect portfolio performance? *The Journal of Investing, 4,* 64-69.

Diltz, J.D. (1995b) The private cost of socially responsible investing. *Applied Financial Economics, 5,* 69-77.

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

Epstein, M.J. (2004) The identification, measurement, and reporting of corporate social impacts: Past, present, and future. *Advances in Environmental Accounting and Management, 2,* 1-29.

Ethix SRI Advisors (2011) *Defining controversial weapons for European institutional investors.* Retrieved from http://www.ethix.se/sites/default/files/Defining%20controversial%20weapons%20for%20investors_Ethix%20SRI%20Advisors%2020110321.pdf

Fabozzi, F.J., Ma, K.C. & Oliphant, B.J. (2008) Sin stock returns. *The Journal of Portfolio Management, 35,* 82-94.

Fama, E.F. & French, K.R. (1993) The cross-section of expected stock returns. *The Journal of Finance, 47,* 427-465.

Freeman, R.E. (1984) *Strategic management: A stakeholder approach.* Boston, MA: Pitman.

Friedman, M. (1962) *Capitalism and freedom.* Chicago, IL: Chicago University Press.

Geczy, C., Stambaugh, R.F. & Levin, D. (2003) *Investing in socially responsible mutual funds.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=416380

Guerard, J.B. (1997) Is there a cost to being socially responsible in investing? *The Journal of Investing, 6,* 11-18.

References

Grinblatt, M. & Titman, S. (1992) The persistence of mutual fund performance. *Journal of Finance, 47,* 1977-1984.

Hamilton, S., Jo, H. & Statman, M. (1993) Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal, 49,* 62-66.

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

Hallahan, T.A. & Faff, R.W. (2010) Induced persistence of reversals in fund performance? The effect of survivorship bias. *Applied Financial Economics, 11,* 119-126.

Jansson, M. & Biel, A. (2011) Motives to engage in sustainable investment: A comparison between institutional and private investors. *Sustainable Development, 19,* 135-142.

Jensen, M.C. (2001) Value maximization, stakeholder theory, and the corporate objective function. *Journal of Applied Corporate Finance, 14,* 8-22.

Keller, G. & Warrack, B. (2003) *Statistics: For management and economics.* Pacifc Grove, CA: Brooks/Cole.

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

Kenneth R. French Data Library (2012) *R_m-R_f, SMB, HML, MOM research factors*. Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

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

KLD Research & Analystics (2006) *Getting started with KLD STATS and KLD's ratings definitions.* Retrieved from http://wrds-web.wharton.upenn.edu/wrds/

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

References

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

Markowitz, H. (1959) *Portfolio selection: Efficient diversification of investments.* New York, NY: Wiley.

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

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

MSCI ESG STATS (2011) *User guide & ESG ratings definition.* Retrieved from http://wrds-web.wharton.upenn.edu/wrds/

Nieuwenhuis, G. (2009) *Statistical methods for business and economics.* New York, NY: Wiley.

MSCI (2012) *Returns series for MSCI KLD 400 Social index, 2004-2011.* Retrieved from http://www.msci.com/products/indices/esg/socially_responsible/performance.html

Orlitzky, M., Schmidt, F.L. & Rynes, S.L. (2003) Corporate social and financial performance: A meta-analysis. *Organization Studies, 24,* 403-441.

Remenyi, D., Williams, B., Money, A. & Swartz, E. (1998) *Doing research in business and management: An introduction to process and method.* London, England: Sage.

Renneboog, L., Ter Horst, J. & Zhang, C. (2008) Socially responsible investment: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance, 32,* 1723-1742.

Rudd, A. (1981) Social responsibility and portfolio performance. *California Management Review, 23,* 55-61.

Salaber, J.M. (2007) *The determinants of sin stock returns: Evidence of the European market.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1071746

SAM (2011) *Alpha from sustainability: SAM white paper.* Retrieved from http://www.sam-group.com/images/Alpha_from_Sustainability_e_tcm794-269011.pdf

References

SAM (2012) *Dow Jones Sustainability Indexes.* Retrieved from http://www.sustainability-indexes.com/dow-jones-sustainability-indexes/index.jsp

Schueth, S. (2003) Socially responsible investing in the United States. *Journal of Business Ethics, 43,* 189-194.

Sekaran, U. (2003) *Research methods for business* (4th ed.). Hoboken, NJ: John Wiley & Sons.

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

Statman, M. (2000) Socially responsible mutual funds. *Financial Analysts Journal, 56,* 30-39.

Statman, M. & Glushkov, D. (2009) The wages of social responsibility*. Financial Analysts Journal, 65,* 33-46.

UN PRI (2012) *An investor initiative in partnership with UNEP Finance Initiative and the UN Global Compact.* Retrieved from http://www.unpri.org

US SIF (2010) *2010 Report on socially responsible investing trends in the United States.* Retrieved from http://ussif.org/resources/pubs/

US SIF (2012) *Performance and SRI Investments.* Retrieved from http://ussif.org/resources/performance.cfm

Vallentin, S. (2003) Socially responsible investing, approaches and perspectives. In M. Morsing & C. Thyssen (Ed.), *Corporate Values and Responsibility,* (pp. 257-267)

WRDS (2012) *Wharton Research Data Services.* Retrieved from http://wrds-web.wharton.upenn.edu/wrds/

Yahoo! Finance (2012) *Returns series for S&P500 and Russell3000 indices, 2004-2011.* Retrieved from http://finance.yahoo.com

Appendices

Table I. KLD/MSCI ESG STATS History

Community (14)		Corporate Governance (16)		Diversity (15)	
Strengths	*Concerns*	*Strengths*	*Concerns*	*Strengths*	*Concerns*
-Charitable Giving (from 1991)	-Investment Controversies (1991 to 2009)	-Limited Compensation (1991 to 2009)	-High Compensation (1991 to 2009)	-CEO (1991 to 2009)	-Controversies (from 1991)
-Innovative Giving (from 1991)	-Negative Economic Impact (from 1991)	-Ownership Strength (1991 to 2009)	-Ownership Concern (1991 to 2009)	-Promotion (from 1991)	-Non-Representation (from 1993)
-Support for Housing (1991 to 2009)	-Tax Disputes (1991 to 2009)	-Transparency Strength (from 1996)	-Accounting Concern (2005 to 2009)	-Board of Directors (from 1991)	-Board Diversity (from 1991)
-Support for Education (1994 to 2009)	-Other Concerns (1991 to 2009)	-Political Accountability Strength (2005 to 2009)	-Transparency Concern (from 2005)	-Work-Life Benefits (from 1991)	-Other Concerns (1991 to 2009)
-Non-US Charitable Giving (1994-2009)		-Public Policy Strength (from 2007)	-Political Accountability Concern (2005 to 2007)	-Women and Minority Contracting (from 1991)	
-Volunteer Programs (2005 to 2009)		-Other Strengths (1991 to 2009)	-Public Policy Concern (from 2007)	-Employment of the Disabled (1991 to 2009)	
-Community Engagement (from 2010)			-Governance Structures Controversies (from 2010)	-Gay and Lesbian Policies (from 1995)	
-Other Strengths (from 1991)			-Other Concerns (from 1992)	-Employment of Underrepresented Groups (from 2010)	
				-Other Strengths (from 1991)	
Total number of Community Strengths (8)	**Total Number of Community Concerns (4)**	**Total Number of Corporate Governance Strengths (6)**	**Total Number of Corporate Governance Concerns (8)**	**Total Number of Diversity Strengths (9)**	**Total Number of Diversity Concerns (4)**

Employee Relations (15/16)		Environment (18/19)		Human Rights (10/14)	
Strengths	*Concerns*	*Strengths*	*Concerns*	*Strengths*	*Concerns*
-Union Relations (from 1991)	-Union Relations (from 1991)	-Beneficial Products and Services (from 1991)	-Hazardous Waste (1991 to 2009)	~~-Positive Record in S. Africa (1994 to 1995)~~	~~-South Africa (1991 to 1994)~~
~~-No-Layoff Policy (1991 to 1993)~~	-Health and Safety Concern (from 1991)	-Pollution Prevention (from 1991)	-Regulatory Problems (from 1991)	-Indigenous Peoples Relations Strength (from 2000)	~~-Northern Ireland (1991 to 1994)~~
-Cash Profit Sharing (from 1991)	-Workforce Reductions (1991 to 2009)	-Recycling (from 1991)	-Ozone Depleting Chemicals (1991 to 2009)	-Labour Rights Strength (2002 to 2009)	-Burma Concern (from 1994)
-Employee Involvement (from 1991)	-Retirement Benefits Concern (1992 to 2009)	-Clean Energy (from 1991)	-Substantial Emissions (from 1991)	-Other Strengths (from 1994)	~~-Mexico (1994 to 2001)~~
-Retirement Benefits Strength (1991 to 2009)	-Supply Chain Controversies (from 1998)	~~-Property, Plant, Equipment (1991 to 1995)~~	-Agriculture Chemicals (1991 to 2009)		-Labor Rights Concern (1998 to 2009)
-Health and Safety Strength (from 2003)	-Other Concerns (from 1991)	-Management Systems Strength (from 2006)	-Climate Change (from 1999)		-Indigenous Peoples Relations Concern (2000 to 2009)
-Supply Chain Policies, Programs & Initiatives (from 2002)		-Other Strengths (from 1991)	-Negative Impact of Products and Services (from 2010)		-Operations in Sudan (from 2010)
-Other Strengths (from 1991)			-Land Use & Biodiversity (from 2010)		-Other Concerns (from 1994)
			-Non Carbon Releases (from 2010)		
			-Other Concerns (from 1991)		
Total Number of Employee Relations Strengths (7)	**Total Number of Employee Relations Concerns (6)**	**Total Number of Environment Strengths (6)**	**Total Number of Environment Concerns (10)**	**Total Number of Human Rights Strengths (3)**	**Total Number of Human Rights Concerns (5)**

Product (11)	
Strengths	*Concerns*
-Quality (from 1991)	-Product Safety (from 1991)
-R+D, Innovation (1991 to 2009)	-Marketing-Contracting Concern (from 1991)
-Benefits to Economically Disadvantaged (from 1991)	-Antitrust (from 1991)
-Access to Capital (from 1991)	-Other Concerns (from 1991)
-Other Strengths (1991 to 2009)	
Total Number of Product Strengths (5)	**Total Number of Product Concerns (4)**

Data: KLD/MSCI ESG STATS (2012)
Note: The numbers between parentheses show the number of strengths and weaknesses incorporated in the ESG ratings data for this thesis.

Table II. Ten industry definitions by the Kenneth R. French Data Library. Each NYSE, AMEX, and NASDAQ stock is assigned to an industry based on its four-digit standard industry classification (SIC) code. On a yearly basis, at the end of June, the companies are designated to a particular industry.

Consumer Non-Durables	Consumer Durables	Manufacturing	Oil, Gas and Coal Extraction	Business Equipment
0100-0999	2500-2519	2520-2589	1200-1399	3570-3579
2000-2399	2590-2599	2600-2699	2900-2999	3622-3622
2700-2749	3630-3659	2750-2769		3660-3692
2770-2799	3710-3711	2800-2829		3694-3699
3100-3199	3714-3714	2840-2899		3810-3839
3940-3989	3716-3716	3000-3099		7370-7372
	3750-3751	3200-3569		7373-7373
	3792-3792	3580-3621		7374-7374
	3900-3939	3623-3629		7375-7375
	3990-3999	3700-3709		7376-7376
		3712-3713		7377-7377
		3715-3715		7378-7378
		3717-3749		7379-7379
		3752-3791		7391-7391
		3793-3799		8730-8734
		3860-3899		

Telephone and Television Transmission	Wholesale, Retail, and Services	Healthcare, Medical Equipment, and Drugs	Utilities	Other – Construction, Hotels, Entertainment, Finance
4800-4899	5000-5999	2830-2839	4900-4949	all other
	7200-7299	3693-3693		
	7600-7699	3840-3859		
		8000-8099		

Data: Kenneth R. French Data Library (2012)

Appendices

Table III. Regression outputs (CAPM, Fama-French three-factor and Carhart four-factor models) of the low-rated portfolio invested in 'shunned' stocks.

Model Summary (CAPM)	
Regression data	
R	0.95
R Square	0.90
Adjusted R Square	0.90
Std. Error of the Estimate	0.02
Observations	96

ANOVA

	df	*Sum of Squares*	*Mean Square*	*F*	*Sig.*
Regression	1	0.32633378	0.32633378	838.7114194	1.24643E-48
Residual	94	0.03657441	0.000389089		
Total	95	0.36290819			

	Coefficients	*Std. Error*	*t*	*Sig.*
Intercept	0.0034	0.002018232	1.677157916	0.10
MKT	1.23	0.042393934	28.96051483	0.00
Annualized return	4.06%			

Model Summary (Fama-French three-factor model)	
Regression data	
R	0.97
R Square	0.93
Adjusted R Square	0.93
Std. Error of the Estimate	0.02
Observations	96

ANOVA

	df	*Sum of Squares*	*Mean Square*	*F*	*Sig.*
Regression	3	0.338615011	0.11287167	427.4530437	7.14931E-54
Residual	92	0.02429318	0.000264056		
Total	95	0.36290819			

	Coefficients	*Std. Error*	*t*	*Sig.*
Intercept	0.0026	0.001666342	1.580547024	0.12
MKT	1.10	0.041611614	26.32970201	0.00
SMB	0.54	0.080139889	6.780808777	0.00
HML	0.03	0.07091106	0.421302614	0.67
Annualized return	3.16%			

Model Summary (Carhart four-factor model)	
Regression data	
R	0.98
R Square	0.95
Adjusted R Square	0.95
Std. Error of the Estimate	0.01
Observations	96

ANOVA

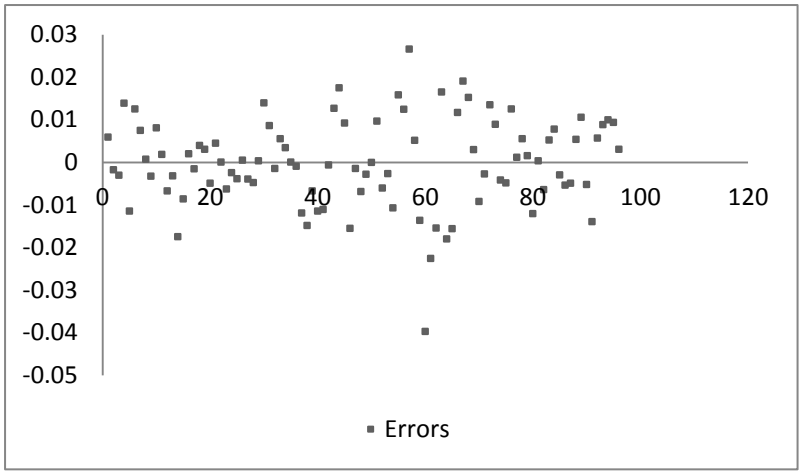
	df	*Sum of Squares*	*Mean Square*	*F*	*Sig.*
Regression	4	0.346040831	0.086510208	466.7256168	1.01655E-59
Residual	91	0.01686736	0.000185356		
Total	95	0.36290819			

	Coefficients	*Std. Error*	*t*	*Sig.*
Intercept	0.0028	0.001396362	2.007132012	0.05
MKT	1.04	0.03609372	28.71637893	0.00
SMB	0.59	0.067588637	8.765325029	0.00
HML	-0.07	0.061637643	-1.201113146	0.23
MOM	-0.19	0.029499919	-6.329499242	0.00
Annualized return	3.36%			

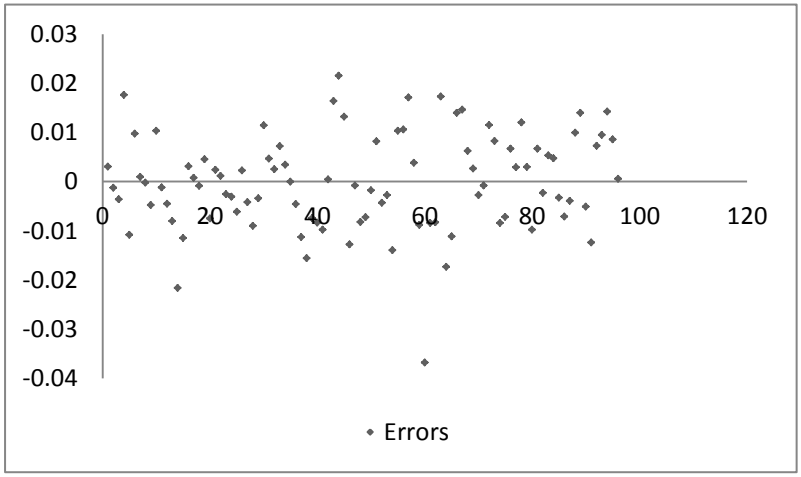
Appendices

Figure IV. Scatter plots of error terms for 'best-in-class (33%)' long-short regression, 2004-2011

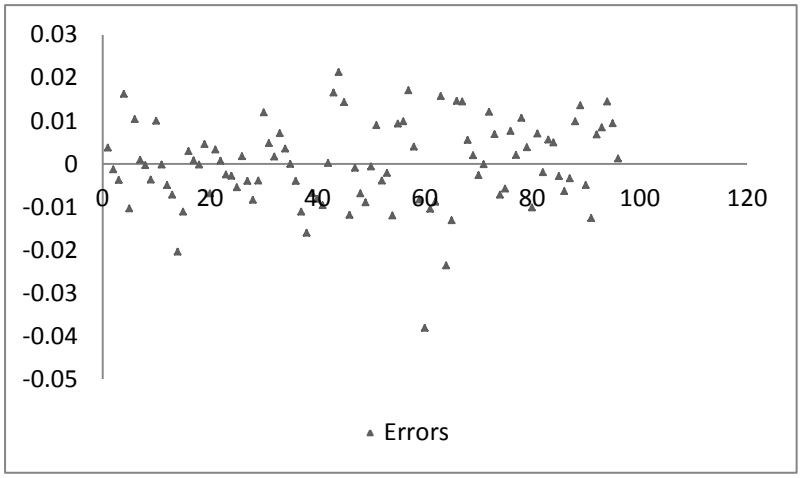
CAPM



Fama-French three-factor model



Carhart four-factor model



Note: The X-axis denote the number of monthly portfolio return observations.
Data: KLD/MSCI ESG STATS (2012)

Table V. Summary of the regression output for '*community (5%)'*. Two long-only portfolios of respectively high- and low-rated stocks, and a third long-short portfolio strategy of buying high-rated stocks and selling low-rated stocks are shown. All portfolios are equally-weighted with ***cut-off points of 5 percent***. For the CAPM, Fama-French three-factor model and Carhart four-factor model the annualized abnormal return (intercept), factor loadings, and the adjusted R Squares over the period 2004-2011 are tabulated. ***, **, and * indicate significance at the 0.01, 0.05, and 0.10 level (*p*-values within parentheses).

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Community top-rated	CAPM	1.26% (0.65)	1.13*** (0.00)				0.85
	3-Factor	0.50% (0.82)	0.95*** (0.00)	0.32*** (0.00)	0.52*** (0.00)		0.91
	4-Factor	0.73% (0.67)	0.88*** (0.00)	0.38*** (0.00)	0.40*** (0.00)	-0.22*** (0.00)	0.94
Community bottom-rated	CAPM	3.35% (0.12)	1.13*** (0.00)				0.91
	3-Factor	2.82% (0.16)	1.04*** (0.00)	0.29*** (0.00)	0.12*** (0.09)		0.92
	4-Factor	2.98% (0.11)	1.00*** (0.00)	0.32*** (0.00)	0.04 (0.52)	-0.14*** (0.00)	0.93
Top- minus bottom-rated (long-short)	CAPM	-2.09% (0.47)	-0.01 (0.86)				-0.01
	3-Factor	-2.33% (0.38)	-0.09* (0.10)	0.03 (0.76)	0.40*** (0.00)		0.14
	4-Factor	-2.25% (0.40)	-0.12** (0.04)	0.05 (0.63)	0.36*** (0.00)	-0.07 (0.12)	0.15

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Table VI. Summary of the regression output for '*employee relations (5%)'*. See table V for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Employee relations *top-rated*	CAPM	1.75% (0.39)	1.18*** (0.00)				0.92
	3-Factor	1.11% (0.53)	1.06*** (0.00)	0.34*** (0.00)	0.20*** (0.00)		0.94
	4-Factor	1.30%	1.01*** (0.00)	0.38*** (0.00)	0.10* (0.06)	-0.17*** (0.00)	0.96
Employee relations *bottom-rated*	CAPM	2.35% (0.40)	1.34*** (0.00)				0.89
	3-Factor	1.09% (0.58)	1.14*** (0.00)	0.73*** (0.00)	0.14** (0.04)		0.94
	4-Factor	1.29% (0.44)	1.08*** (0.00)	0.78*** (0.00)	0.04 (0.49)	-0.18*** (0.00)	0.96
Top- minus bottom-rated (long-short)	CAPM	-0.60% (0.80)	-0.16*** (0.00)				0.12
	3-Factor	0.03% (0.99)	-0.08* (0.09)	-0.40*** (0.00)	0.05 (0.50)		0.27
	4-Factor	0.01% (1.00)	-0.07 (0.13)	-0.40*** (0.00)	0.06 (0.46)	0.01 (0.72)	0.26

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

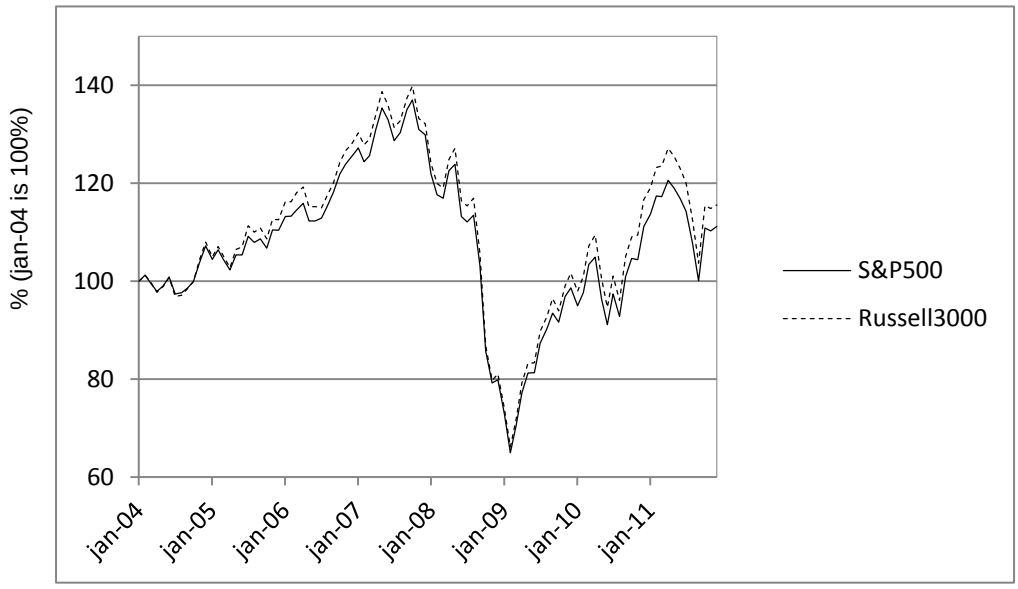
Appendices

Table VII. Summary of the regression output for '*product (5%)'*. See table V for description.

Portfolio	Performance Benchmark	Annualized Excess Returns	Market Factor	Small-Large Factor	Value-Growth Factor	Momentum Factor	Adjusted R^2
Product top-rated	CAPM	3.15%	1.30***				0.89
		(0.25)	(0.00)				
	3-Factor	1.97%	1.11***	0.67***	0.18**		0.94
		(0.33)	(0.00)	(0.00)	(0.01)		
	4-Factor	2.14%	1.05***	0.71***	0.10	-0.16***	0.95
		(0.23)	(0.00)	(0.00)	(0.15)	(0.00)	
Product bottom-rated	CAPM	2.78%	1.11***				0.93
		(0.12)	(0.00)				
	3-Factor	2.44%	1.02***	0.12**	0.31***		0.95
		(0.11)	(0.00)	(0.05)	(0.00)		
	4-Factor	2.62%**	0.97***	0.16***	0.22***	-0.16***	0.97
		(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	
Top- minus bottom-rated (long-short)	CAPM	0.36%	0.18***				0.13
		(0.89)	(0.00)				
	3-Factor	-0.47%	0.08*	0.55***	-0.13		0.37
		(0.84)	(0.09)	(0.00)	(0.12)		
	4-Factor	-0.47%	0.08	0.55***	-0.13	0.00	0.36
		(0.84)	(0.10)	(0.00)	(0.14)	(0.96)	

Data: MSCI ESG STATS (2012)
Adapted from: Kempf & Osthoff (2007) Statman & Glushkov (2009)

Figure VIII. Performance of the MSCI ESG STATS rated indices, 2004-2011



Data: Yahoo! Finance (2012)