The Price of Ethical Investing:

Evaluating the performance of socially responsible indices

Martin H. Bredal **Nicolai Negård**

Supervisor: Thore Johnsen

Master thesis in Financial Economics

NORWEGIAN SCHOOL OF ECONOMICS
Bergen
Spring, 2015

 We would like to thank our supervisor Thore Johnsen for excellent guiding.

Abstract

We use a Fama-French based approach to investigate the risk-adjusted performance of five regionally diverse sets of SRI indices and their conventional benchmarks from 1997 to 2014. In accordance with most previous research, we find that SRI indices perform on par with their benchmarks in the long run. However, we postulate that SRI screening leads to increased idiosyncratic risk and that this will translate into inferior risk-adjusted returns in periods of falling markets. Expanding on the Fama-French approach with dummy variables for the Dotcom Fall in the early 2000s and the Financial Crisis of 2007 to 2009, as well as adjusting for variations in the market, size and value premiums in these periods, we find that SRI underperforms in periods of falling markets. As a result, we argue that socially responsible investors with a long investment horizon should not expect inferior financial returns, but investors with a shorter investment horizon should be wary of SRI.

Contents

1 Introduction

Socially Responsible Investing (SRI) works by incorporating non-financial information, such as the Environmental, Social and Governance (ESG) behavior of companies, into the investment decision. A social responsible investor can avoid firms or industries involved in ethically questionable activities, or he can take a more proactive approach and seek out firms or industries with impeccable ethical records. Either way, this reduces the available investment universe, which modern portfolio theory says will lead to increased idiosyncratic risk and less efficient portfolios (Humphrey and Tan 2014). It is a stylized fact of finance that investors in efficient capital markets do not receive compensation for taking on idiosyncratic risk, which implies that SRI should underperform conventional investments on a risk-adjusted basis. However, supporters of SRI argue that the benefits of integrating ESG considerations into the portfolio creation process exceed the loss in efficiency caused by the reduced investment universe (RBC 2012). Most research on SRI performance fail to reject the null hypothesis that SRI performs on par with conventional investments in the long run, supporting neither of these arguments (Kurtz and DiBartolomeo 2011, Boon et al. 2013, Humphrey and Tan 2014). For long-term investors, this might provide a satisfactory answer regarding the performance of SRI versus conventional investing. On the other hand, investors with a shorter investment horizon might shun SRI if it leads to an inferior risk-adjusted return in the short term. Considering the reduced diversification suggested above, this seems entirely plausible. If SRI delivers an inferior risk-adjusted return in the short term, we suspect that this will become apparent in times of falling markets, when having a well-diversified portfolio is crucial. Therefore, this study investigate both the long-term risk-adjusted return of SRI and the short-term performance in periods of falling markets.

Most research on SRI performance investigate SRI funds, but this introduces distorting effects such as market timing, manager skill and management fees (RBC 2012). Instead, this study expands on previous research by Gjølberg and Johnsen (2008) on the risk-adjusted performance of SRI indices. We investigate five regionally diverse sets of SRI indices and their benchmarks over an 18-year period from 1997 to 2014. Through a Fama-French three-factor model, we account for differences in systematic risk factors between the SRI indices and their benchmarks, specifically the tendency of SRI to tilt towards large growth stocks (Boon et al. 2013, Renneboog et al. 2008, Gjølberg and Johnsen 2008). To measure the SRI performance in periods of falling markets, we expand on the Fama-French three-factor model with dummy variables for the Dotcom Fall in the early 2000s and the Financial Crisis of 2007 to 2009. Furthermore, to adjust for the large variations in the market, size and value premiums in these periods, we include interaction terms between the dummy variables and the Fama-French risk factors.

In accordance with most previous research, the results show no significant difference in the performance of the SRI indices and their benchmarks over the full sample period. This implies that SRI might be a suitable investment option for investors with an infinite investment horizon, such as pension funds and university endowment funds. However, the results indicate substantial financial losses from investing in the SRI indices instead of their benchmarks in both periods of falling markets. For example, FTSE4Good US, a well-known SRI index, deliver an inferior risk-adjusted return compared to its benchmark of -6.59 %-points p.a. during the Financial Crisis. This inferior risk-adjusted return in periods of falling markets should make investors with a shorter investment horizon wary of SRI.

We organize the remainder of this study as follows: Section 2 presents an overview of the SRI industry today and its different approaches, while section 3 discusses theory. Section 4 reviews related literature, section 5 describes the data used and section 6 presents descriptive statistics. Section 7 introduces the models, while section 8 presents the results. Section 9 provides a discussion of those results and section 10 concludes.

2 Socially Responsible Investing

SRI has many definitions, but we define it as investments that are limited to some degree by environmental, social or governance criteria. This definition is consistent with that of Schröder (2007) and Renneboog et al. (2008). SRI is tightly related to Corporate Social Responsibility (CSR) and can be seen as a component of CSR overall. The EU Commission (2011) define CSR as ”companies taking responsibility for their impact on society”.

2.1 The SRI Industry

SRI is a fast growing industry, especially in the U.S. and Europe. At the end of 2014, $6.57 trillion were managed using SRI strategies in the U.S.. This equals one out of every six dollars under professional management and is an increase of 76 % from $3.74 trillion in 2012. In Europe, the total size of the SRI industry is larger than in the U.S., but the growth rate is lower. €9.9 trillion of assets were managed with SRI strategies in 2013, which is an increase of 22 % from 2011 (EUROSIF 2015). The increased popularity of SRI has translated into a large growth of investor services trying to meet this demand. For example, U.S. SRI funds have grown from $12 billion and 55 funds in 1995 to $4.3 trillion and 925 funds in 2014 (USSIF 2014). Furthermore, index producers like FTSE and MSCI have designed a wide range of indices with different SRI strategies, while companies like Ethics and EIRIS have specialized in research and advisory services on SRI. In addition, established assurance and advisory firms, such as EY and PwC have started their own departments for advisory within climate, sustainability and social responsibility.

2.2 Classification of SRI approaches

European Sustainability Investment Forum (Eurosif) presents seven distinct strategies for socially responsible investing: exclusions, norms-based screening, best-in-class selection, sustainability themed, ESG integration, engagement and impact investing (EUROSIF 2015). We group these investment strategies into three main bodies, namely Negative screening, Positive screening and Engagement.

2.2.1 Negative screening

With negative screening, the investor excludes certain industries or firms in which he does not wish to be involved. This strategy has gone mainstream and now covers more assets than any other SRI approach. In Europe alone, exclusions cover about 41 % of professionally managed assets, with the most common exclusions covering cluster munition and anti-personnel landmines (EUROSIF 2015). Norms-based screening is another type of negative screening, which involves applying an ESG-filter to the portfolio manager's investment universe. These filters are created by index management firms or ethical advisory firms, and they are often based on international norms such as the United Nations Global Compact or the OECD Guidelines for Multinational Corporations and International Treaties (EUROSIF 2015).

2.2.2 Positive screening

Positive screening is an umbrella term covering many different SRI strategies. However, all the strategies have in common that they seek out firms with a proactive approach to ESG issues. Best-in-class selection involves investing in the best companies in an industry according to ESG criteria. ESG integration involves a systematic approach of including ESG risks and opportunities into the financial analysis and investment decisions. Sustainability themed investing are typically direct investments into industries transitioning into more sustainable forms of production and energy consumption, while impact investing includes a social goal in addition to a financial goal. For example, microfinance is an impact investing method that attempts to further social integration while creating competitive financial returns.

2.2.3 Engagement

Engagement involves taking on the role of active ownership, either by using shareholder votes or through communication with the company, to improve the environmental, social or governance aspect of the company. On the index level, it revolves around providing firms that fall beneath a certain threshold on ESG criteria with a warning before excluding them, making it possible for these firms to correct their behavior (FTSE 2014c).

3 Theory

3.1 Why investigate market-level data

There is an extensive literature attempting to clarify the relationship between corporate social performance (CSP) and corporate financial performance (CFP) on the firm level. Supporters of a positive relationship typically use stakeholder theory to argue that firms engaging in ESG activities outperform their competitors (RBC 2012). Proponents of a negative relationship, most famously Nobel Prize winner Milton Friedman, argues that corporate executives' only responsibility is to their shareholders and that CSR practices goes against this (Friedman 1970). When it comes to studying CSP and its effect on CFP at the firm level, problems of determining the direction of causality arises. Does good corporate social behavior lead to good financial performance or does good financial performance simply allow a firm to engage in social practices? This field of study has largely produced inconclusive results. An extensive meta study by Margolis et al. (2007) investigate 167 CSP/CFP studies only to find a weak positive correlation. They conclude that further studies of CSP/CFP at the firm level is of little use and discourage further effort into this area. A more worthwhile approach may be to investigate market-level data.

3.2 Index approach

The most widely used approach for investigating SRI performance through market-level data is to compare the performance of SRI funds to conventional funds. However, investigating fund performance introduces distorting effects such as market timing, management skills and transaction costs. Furthermore, fund data often suffers from survivorship bias (Carhart 1997) and presents a problem of finding matching funds to use as benchmarks (Boon et al. 2013). To avoid these problems, we prefer to investigate the performance of SRI indices instead. For example, most SRI indices are screened versions of conventional indices, circumventing the problem of finding an appropriate benchmark. However, investigating the performance of SRI indices versus their benchmarks is not problem free. For instance, style differences between the SRI indices and their conventional benchmarks, such as different loadings on the size and the value factor, will have to be adjusted for (Gjølberg and Johnsen 2008, RBC 2012).

3.3 Implications of SRI screening in portfolio management

3.3.1 Underperformance hypothesis - *"Doing good, but not well"*

Imposing a negative screen on a portfolio reduces the available investment universe. Furthermore, if this negative screen affects different industries unevenly, this might lead to skewed sector weights. Proponents of the underperformance hypothesis advocate that reducing the investment universe and changing the sector tilts will impede the portfolio manager's ability to form fully diversified portfolios. This may result in portfolios with increased idiosyncratic risk. In efficient capital markets, investors do not receive higher return as compensation for taking on idiosyncratic risk. Therefore, modern portfolio theory suggests that imposing negative screens should lead to less efficient portfolios with lower risk-adjusted returns (Humphrey and Tan 2014). Additionally, if ESG factors are negatively related to financial performance, then screening for these factors may cause the portfolio to underperform. For example, Hong and Kacperczyk (2009) study the performance of publicly traded companies involved in tobacco, gaming and alcohol, so called "sin" stocks, and find that these consistently outperform comparable stocks. They attribute this outperformance to sin stocks being less followed by Wall Street analysts and less held by norm-constrained institutions, such as pension funds.

3.3.2 Outperformance hypothesis - *"Doing well by doing good"*

Supporters of SRI argue that the benefits of integrating ESG considerations into the portfolio creation process outweighs the loss of portfolio efficiency caused by the reduced investment universe. They maintain that companies excluded by SRI screening are involved in unsustainable activities that makes them less profitable over time. For example, heavy polluters are more likely to face litigation (RBC 2012). If market participants systematically underestimate the benefits or overestimate the costs of being socially responsible, then the expected return of socially responsible companies might be consistently higher (RBC 2012, Statman and Glushkov 2009). This is consistent with the findings of Edmans (2011) who shows that firms with greater employment practices consistently deliver superior earnings performance, but that the market undervalues these intangibles. Furthermore, Derwall et al. (2005) show that companies with high environmental records outperform comparable stocks. It is also in agreement with Kempf and Osthoff (2007) who finds that stocks of firms ranking high on human rights, employee relations, environment, community, diversity and products outperform low ranking stocks.

3.4 Long-term versus short-term SRI performance

Many of the arguments in favor of the outperformance hypothesis, for example the findings of Edmans (2011) on the systematic failure of the market to value intangibles, are likely to require some time to manifest into positive excess returns. This suggests that we should investigate a long sample period. For institutional investors, such as pension funds, who are mostly interested in potential portfolio shortfall far into the future, investigating a long sample period might provide a satisfactory answer regarding the performance of SRI versus conventional investments. Contrary, private investors might shun investing in SRI if the short-term negative fluctuations are larger than for conventional investments. This may also be the case for institutional investors under governmental control. For example, the Norwegian Government Pension Fund Global is under heavy public scrutiny, and short-term underperformance tend to produce newspaper headlines and political dismay. If the underperformance hypothesis is correct, and the reduced investment universe and changed sector tilts lead to a substantial increase in the idiosyncratic risk, then this is likely to become apparent in times of crisis, when having a diversified portfolio is of the utmost importance. To accommodate the interests of both types of investors, this study tests the effects of SRI screening over a long holding period and during periods of drastically falling stock markets.

4 Literature review

Renneboog et al. (2008) builds on their previous research (Renneboog et al. 2007) and investigate the performance of SRI funds from 17 countries. They find that SRI funds in the U.S., U.K. and in many European and Asia-Pacific countries underperform their benchmarks by -2.2 % to -6.5 % p.a. However, when risk is taken into consideration through a four-factor model, they do not find any statistical evidence that the returns of SRI funds differ from that of conventional ones. Furthermore, they investigate whether increasing the SRI screening intensity enhances fund performance. They find that funds with one additional screen are associated with 1 % p.a. lower factor-adjusted return, and conclude that high screening intensity constrains the risk-return optimization.

Kurtz and DiBartolomeo (2011) performs a holdings-based analysis using a multifactor model composed of beta, industry dummies and fundamental factors to investigate the performance of the KLD400 Social index from 1992 to 2010. They find that most of the outperformance of KLD400 over S&P500 is factor-driven. For example, KLD400's higher beta accounted for 2/3 of the outperformance over the full sample period. They continue by dividing the sample period into two. The period of nominal outperformance by KLD400 up until 1999 ”appears to have been entirely factor-driven, with beta, industries, and fundamental factors accounting for virtually all of the observed active return”. However, in the following period of underperformance until 2010, ”the factor bets are simply reversed”. Overall, they find no significant alpha in either direction. The authors conclude that this is unsurprising, because the research database for KLD400 has been available for quantitative analysts since the early 1990s and that alpha-seeking investors would exploit any ESG-alphas as soon as they were discovered.

Managi et al. (2012) use the Markov Switching model on return data of SRI indices and their benchmarks in the U.S., U.K. and Japan from the early 2000s until 2008. They argue that, ”even if two markets have similar unconditional expected return and volatility, they could be considerably different as conditional on the regime”. Their results show two distinct regimes, bull and bear, for the SRI indices and their benchmarks for all three countries. More importantly, they find that these two regimes coincide in both occurrence and length for the SRI indices and their benchmarks. Furthermore, they find strong evidence of co-movement between the SRI indices and their benchmarks in each regime. Lastly, they find no statistical difference in the mean or volatility of the SRI indices and their benchmarks in either bull or bear regimes. Overall, they conclude that even when looking at conditional mean and volatility, their findings is in alignment with much of the literature, which fails to find a significant underperformance of SRI.

Boon et al. (2013) use a characteristic-based approach to test the performance of SRI funds. This characteristic-based approach disaggregates performance into three components: manager stock selection ability; manager characteristic timing ability and fund style. They find that constraining portfolios, at least to a certain degree, neither enhance nor hinder fund manager's ability to generate returns. Furthermore, they find that SRI managers appear to be better able to time style characteristics, especially the book-to-market factor, than their conventional counterparts. Lastly, they find that SRI funds are biased towards large capitalization stocks, which underperformed over their sample period (2000 to 2010). They argue that this is because large firms are able to devote more resources to meet the ESG demands of positive screening.

Humphrey and Tan (2014) replicate 10,000 pairs of SRI and conventional portfolios to test the impact of SRI screening on performance. Through this process, they remove potentially interfering effects, such as manager skill and transaction costs. Measuring performance through a one- and a four-factor model, they find no significant difference in the risk-adjusted return of screened and unscreened portfolios. They conclude that a typical SRI fund will neither gain nor lose from screening its portfolio. In a preceding working paper using the same methodology, Humphrey and Tan (2011) investigate the different effects of positive and negative screens on the performance of SRI funds. They find that positive screening results in increased return, but also increased total risk and beta. Lastly, they find that increasing the number of negative screens reduces the ability of SRI funds to diversify.

Belghitar et al. (2014) criticize previous SRI research for not testing the prerequisites for the methods used. For example, they argue that much of the previous research on SRI ignores the non-normal distribution of return data. They solve this by using the Marginal Conditional Stochastic Dominance (MCSD) framework, which can accommodate any return distribution. Based on weekly data for six SRI indices, from July 2001 to November 2010, they provide evidence that social responsible investing comes at a financial price. More specifically, they conclude that there is nothing to be gained or lost from SRI in terms of mean and variance, but in the higher moments of the return distribution there is a price to be paid. On average, they find that conventional indices, compared to their SRI counterparts, have a 27 % higher skewness and 15 % lower kurtosis. They conclude that risk-averse investors can gain a higher utility by reducing their socially responsible holdings in favour of conventional ones.

5 Data

Suitable indices should cover a long time span and have proper benchmarks for comparison. We find that the KLD400 Social index from the American index producer MSCI (MSCI 2014) and four FTSE4Good indices from the British index producer FTSE (FTSE 2014b), fulfills these requirements. Table 1 presents the SRI indices and their respective benchmarks, as suggested by the index producers, used in this study. The indices provide us with good coverage of the European and the American SRI market, which are the most developed ones, as well as the global market. The FTSE4Good index family was created in 2001 and FTSE constructed backtracking series back to January 1997, providing us with 18 years of data for these indices. KLD400 was founded in 1990, but to obtain consistency across the data set with regards to time span and sub-periods, we only use data back to January 1997.

Table 1: The SRI indices and their benchmarks

		# of constituents[1]			Screening	
SRI	**Benchmark**	**SRI**	**BM**	**SRI/BM**[2]	**Positive**	**Negative**
FTSE4G US	FTSE USA	174	656	27 %	Yes	TAN
FTSE4G UK	FTSE All-Share	243	642	38 %	Yes	TAN
FTSE4G Europe	FTSE Dev Europe	312	519	60 %	Yes	TAN
FTSE4G Global	FTSE Dev World	780	2115	37 %	Yes	TAN
KLD400	MSCI USA	400	628	64 %	Yes	AGTAFA + GMO

[1] Number of constituents per December 31st 2014.
[2] Number of constituents in the SRI indices divided by the number of constituents in their respective benchmarks.
TAN = Tobacco, Armaments, Nuclear power/weapons.
AGTAFA = Alcohol, Gambling, Tobacco, Armaments, Firearms, Adult entertainment.
GMO = Genetically Modified Organisms.
Table inspired by Gjølberg and Johnsen (2008).

5.1 MSCI KLD400

MSCI KLD400 is one of the oldest and best known of the socially responsible indices. It was launched May 1st 1990 by KLD as the Domini 400 index and has had multiple third-party index administrators since. MSCI assumed administration of the KLD400 in September 2010. The screening of the index is conducted using data from MSCI ESG Research. It has an AGTAFA + GMO negative screening, meaning companies within alcohol, gambling, tobacco, armaments, firearms, adult entertainment and genetically modified organisms are excluded. MSCI combines the negative screening with an ESG best-in-class methodology. The index is composed of 400 constituents, and has a target of only 200 large and mid-cap companies. Due to this target, KLD400 consists of a large amount of small-cap companies. Previously, the S&P500 was used as a benchmark for KLD400, but MSCI use their own MSCI USA equity index instead (MSCI 2014). MSCI USA also contains a large amount of small-cap companies, and we therefore agree that MSCI USA is a better benchmark than S&P500, and use the former as a benchmark for KLD400 in this study.

5.2 FTSE4Good

The FTSE4Good indices are screened versions of conventional FTSE indices, covering different regions and countries. The European and global FTSE4Good versions are based on FTSE Developed Europe and FTSE Developed World, respectively. The investable universe of the U.S. version is FTSE USA, while the U.K. version is based on FTSE All-Share, the broad market index of the London Stock Exchange (FTSE 2014b). FTSE uses an extensive ESG-rating model with more than 300 indicators to screen these indices. This model utilizes both positive and negative screens, where the positive screening happens in conjunction with the ethical research firm EIRIS. The negative screening results in an exclusion of tobacco, armaments, nuclear weapons and nuclear power (FTSE 2014c). Table 1 shows that this screening process results in an inclusion of between 27-60 % of the constituents in the benchmark indices.

5.3 Fama-French Factors

The Fama-French model was originally made for the American market, and analysis conducted on other regions had to use the U.S. Fama-French factors. Kenneth French has later added separate factors for developed markets to his database (French 2015). Fama and French (2012) show that these factors are better suited for regional analysis. Therefore, we use separate Fama-French factors for the U.S., European and global markets. It is important to note that all these factors are denominated in U.S. dollars and use the U.S. one-month Treasury bill as the risk-free rate. With dollar denominated Fama-French factors as independent variables we require dollar denominated returns for all the dependent variables as well, i.e. all the indices. Otherwise, exchange rate fluctuations will disrupt the results in the Fama-French regressions. FTSE4Good UK and FTSE4Good Europe are denominated in Pounds and Euros respectively, and must therefore be calculated into U.S. dollars.

5.4 Risk-free rate

When calculating the Sharpe ratios of the indices, we need the risk-free rate of return. For the U.S. and global indices, we use the one-month U.S. Treasury bill. The Sharpe ratios for the U.K. and European indices are calculated using return data denominated in their local currencies and therefore require a risk-free rate that matches. For the U.K., we use the British three-month Treasury bill. Obtaining the risk-free rate for Europe can be difficult, since our sample period spans the introduction of the Euro. A solution is to use the government bill of a Euro-zone member, and we find the German three-month Bubill to be the best choice.

5.5 Data collection

We download monthly total return data from January 1997 to December 2014 in USD from Thomson Reuter's DataStream (2015) for all the indices. For FTSE4Good UK, FTSE All-Share and the European indices, we also gather the monthly total return data in their local currencies. Specifically, we use DataStream's Total Return Index (RI), which reinvests dividends. We download the Fama-French factors from Kenneth French's homepage (French 2015), containing the market, value and size factors as well as the U.S. risk-free rate. The U.K. risk-free rate is retrieved from Bank Of England (2015), while we get the German risk-free rate from Bloomberg.

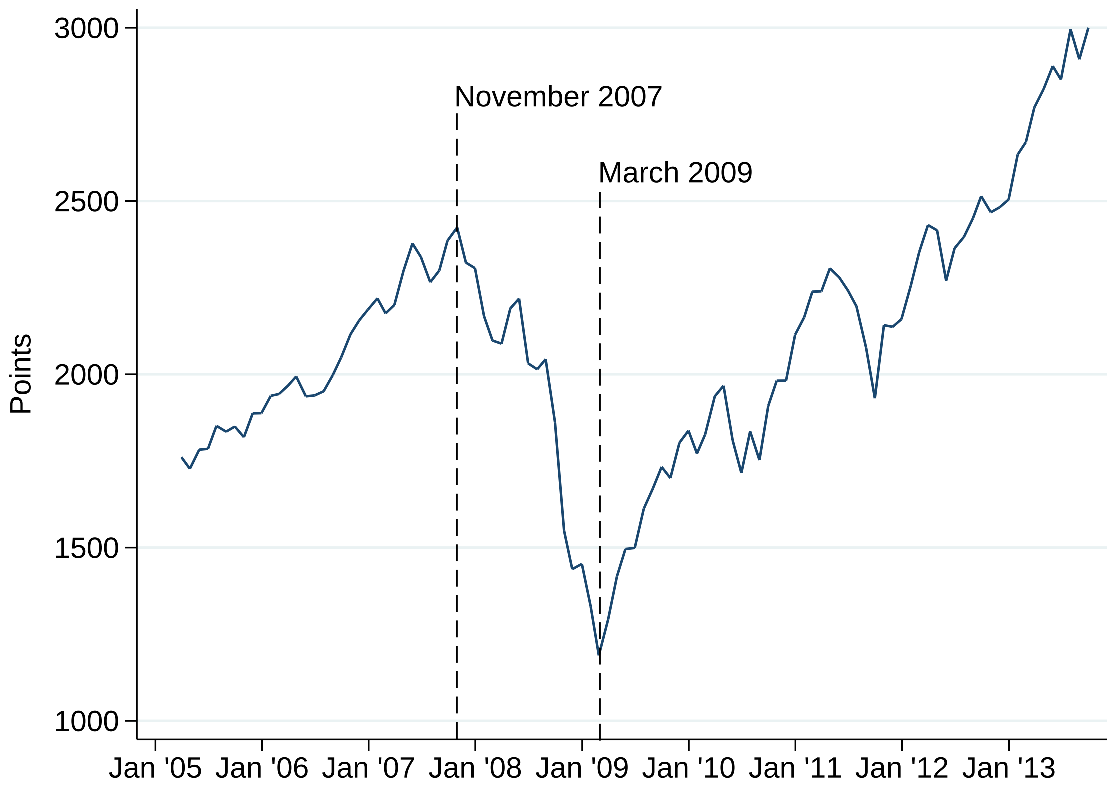
5.6 Sub-periods

The period from January 1997 to December 2014 consists of five distinct periods in the stock markets: The tech-bubble in the late 1990s, the Dotcom Fall in the early 2000s, the growth period in the mid-2000s, the Financial Crisis from 2007 to 2009 and finally the turbulent recovery period from 2009 until today, disrupted by the European sovereign debt crisis.

On March 10th 2000, NASDAQ peaked (Bloomberg), marking the end of the tech-bubble of the late 1990s. The stock markets fell through the spring and the financial climate did not change until April 2003, defining the Dotcom Fall as April 2000 through March 2003. The Financial Crisis can be dated back to August 2007, when BNP Paribas blocked withdrawals from hedge funds because of complete lack of liquidity (Elliot 2012). However, this did not manifest itself in the stock market before November 2007 (Figure 1). It is not clear exactly when the crisis ended, but we see a distinct change in the return of the S&P500 in March 2009, and therefore define the Financial Crisis as November 2007 through February 2009. The following years were plagued by the European sovereign debt crisis, but its effects on the stock markets were drastically smaller and more sporadic than that of the Dotcom

Fall and the Financial Crisis. Combined with the overall upwards trend of the stock markets following the Financial Crisis, we do not find it beneficial to separate the European sovereign debt crisis into a distinct period of falling markets. Hence, the three growth periods in the sample are January 1997 through March 2000, April 2003 through October 2007 and March 2009 until the end of the sample in December 2014.

Figure 1: S&P 500 Total Return



(Bloomberg)

6 Descriptive statistics

6.1 Full sample period: 1997-2014

Table 2 presents the annualized mean return, standard deviation and Sharpe ratios for all the indices over the full sample period. We observe that the difference in mean return between the SRI indices and their benchmarks are small, ranging from –0.43 %-points p.a. for FTSE4Good UK to 0.32 %-points p.a. for KLD400. Furthermore, all SRI indices, except FTSE4Good UK, have a larger volatility than their benchmarks. The resulting Sharpe ratios are smaller for four out of five SRI indices, indicating that SRI underperforms on a risk-adjusted basis over the full sample period.

Table 2: SRI vs Benchmark Jan'97 - Dec'14
% Annualized Mean and Standard deviation

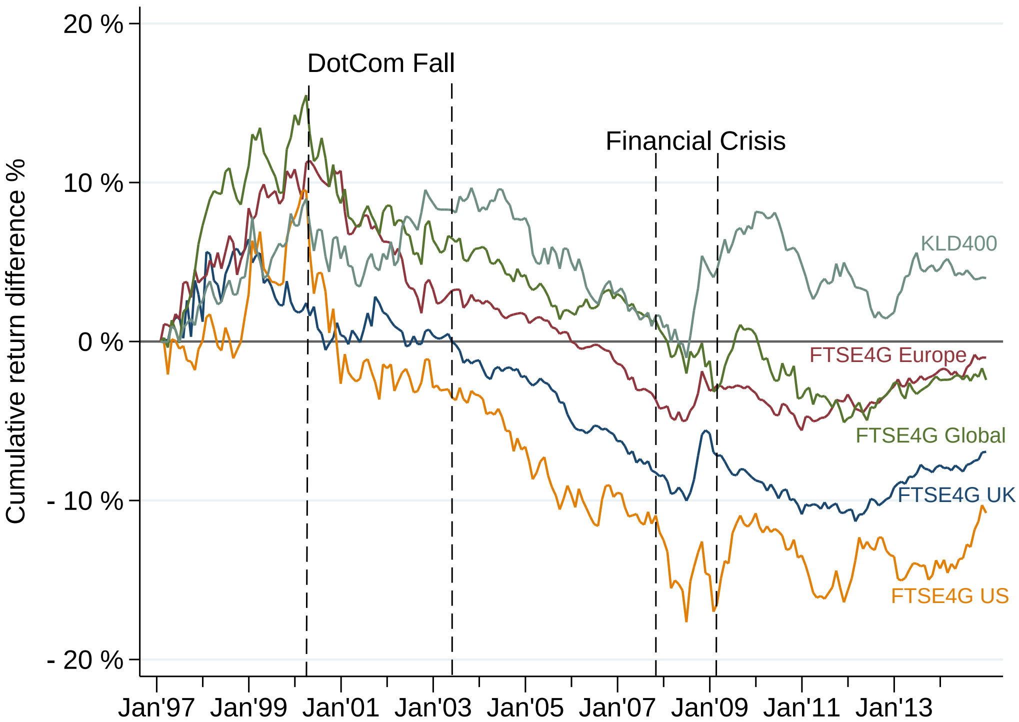
	Full Sample Period		
	Mean return	**St.Dev**	**Sharpe**
FTSE4Good US	9.02	17.12	0.39
FTSE USA	9.31	15.66	0.43
Difference	-0.29	1.46	-0.04
FTSE4Good UK	7.20	13.98	0.27
FTSE All-Share	7.63	14.30	0.29
Difference	-0.43	-0.32	-0.02
FTSE4Good Europe	8.34	16.80	0.37
FTSE Dev Europe	8.38	16.32	0.39
Difference	-0.04	0.48	-0.02
FTSE4Good Global	8.33	17.02	0.35
FTSE Dev World	8.10	15.87	0.36
Difference	0.23	1.15	-0.01
KLD400	9.48	16.03	0.44
MSCI USA	9.16	15.71	0.43
Difference	0.32	0.32	0.01

Mean return is annualized by $(1+R_{month})^{12}-1$. R_{Month} is the arithmetic monthly mean. Stdev is annualized by $StDev_{Month} * \sqrt{12}$. $StDev_{Month}$ is the arithmetic monthly mean. Sharpe ratio = $(\mu_p - r_f)/\sigma_p$

Figure 2 presents the cumulative return difference for the pairs of SRI and conventional indices. Looking at cumulative return differences is useful because it reveals the total excess return from holding an SRI index instead of its benchmark from January 1997. The cumulative return difference is calculated using geometric mean and shows an even grimmer picture of SRI. This can be explained by the large fluctuations in the return data, which causes the geometric mean to be lower than the simple arithmetic mean (Gjølberg and Johnsen 2003). We now observe that four out of five SRI indices deliver inferior returns when held over the full sample period, compared to only three when using arithmetic mean (Table 2). An investor holding the FTSE4Good US index over the full sample period would earn 10.8 %-points lower cumulative return, or -0.63 %-points p.a., than an investor holding FTSE USA over the same period. The only SRI index providing a larger cumulative return than its benchmark over the full sample period is KLD400 at 4.0 %-points, or 0.22 %-points p.a..

Figure 2: Cumulative Return Difference

% Cumulative return difference between SRI and BM.



(DataStream 2015)

6.2 Dotcom Fall & Financial Crisis

Table 3 presents the annualized arithmetic mean return and standard deviation for all the indices under the Dotcom Fall and the Financial Crisis. During the Dotcom Fall, all the SRI indices deliver inferior returns at a higher volatility, except for FTSE4Good UK, which have a marginally lower volatility. The worst performing SRI index in this period is FTSE4Good US, delivering an annualized 3.01 %-points lower return at 2.28 %-points higher volatility than its benchmark. In contrast to the Dotcom Fall, there are no clear patterns in the descriptive statistics for the Financial Crisis. For example, we observe that FTSE4Good US deliver an annualized 2.68 %-points lower return at 1.40 %-points higher volatility than FTSE USA, while KLD400 deliver an annualized 1.18 %-points higher return at 0.68 %-points lower volatility than MSCI USA.

Table 3: SRI vs. Benchmark - Falling stock markets

% Annualized Mean and Standard Deviation

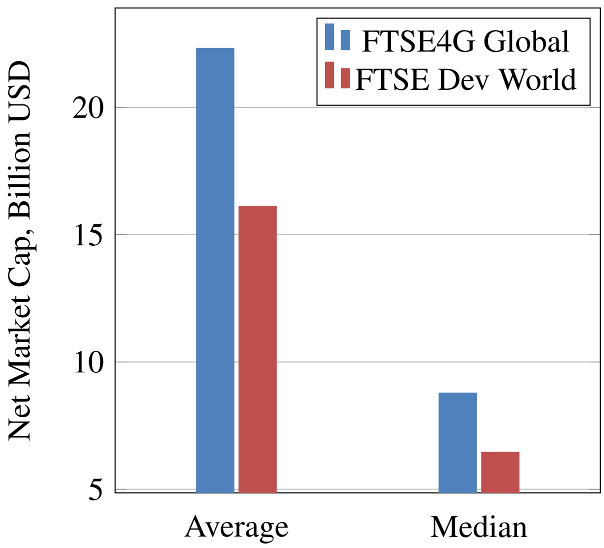
	Dotcom Fall		**Financial Crisis**	
	Mean Return	**StDev**	**Mean Return**	**StDev**
FTSE4G US	-18.15	19.79	-42.39	20.49
FTSE USA	-15.14	17.51	-39.71	19.09
Difference	-3.01	2.28	-2.68	1.40
FTSE4G UK	-14.88	15.50	-30.89	18.81
FTSE All-Share	-14.21	16.04	-31.43	19.09
Difference	-0.67	-0.54	0.54	-0.28
FTSE4G Europe	-21.38	19.30	-42.17	17.93
FTSE Dev Europe	-19.42	18.24	-42.51	18.36
Difference	-1.96	1.06	0.34	-0.43
FTSE4G Global	-18.38	17.87	-44.13	21.51
FTSE Dev World	-16.20	16.07	-42.34	21.00
Difference	-2.18	1.80	-1.79	0.51
KLD400	-16.01	17.90	-38.72	18.48
MSCI USA	-15.92	17.33	-39.90	19.16
Difference	-0.09	0.57	1.18	-0.68

Mean return is annualized by $(1+R_{Month})^{12}-1$. R_{Month} is the arithmetic monthly mean.
Stdev is annualized by $StDev_{Month} * \sqrt{12}$. $StDev_{Month}$ is the arithmetic monthly mean.
All Sharpe ratios for the Dotcom Fall and the Financial Crisis are negative, and therefore excluded because of their potential misleading interpretations (Israelsen 2003). For those interested, the Sharpe ratios can be found in Table A1.

6.3 Size and sector characteristics

An interesting feature of the SRI indices is their tilt towards specific sectors and larger firms (Gjølberg and Johnsen 2008, Boon et al. 2013). Figure 3 illustrates the average and median market capitalization of the constituents for FTSE4Good Global and its benchmark. The average constituent in FTSE4Good Global is 38.1 % larger than the benchmark's, while the median constituent is 36 % larger. Boon et al. (2013) argue that this tilt towards large firms is because they are able to devote more resources to meet the ESG demands of positive screening. Another reason entirely may be that large firms are more in the public's view and may reap a greater reputational benefit from being part of an SRI index, and as a consequence intensifies their CSR practices (Mortier 2014). For example, the long-lasting boycott of Nestlé, the world's biggest producer of infant formula, was lifted by many organizations after its inclusion in the FTSE4Good index family in 2011 (Nestlé 2011a). However, to remain included in the FTSE4Good index family, Nestlé is required to have their infant formula marketing practices continuously verified by PwC (Nestlé 2011b), which is expensive, illustrating that Boon et al. (2013) also have a valid argument.

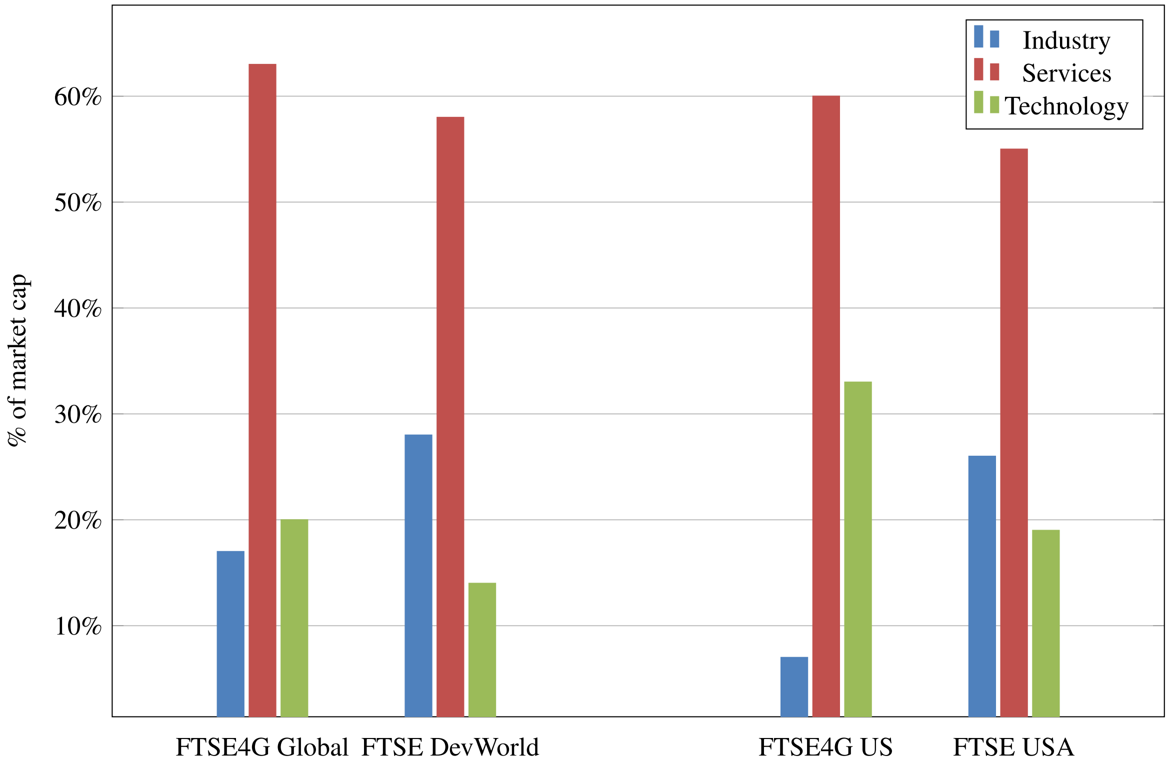
Figure 3: Average and median size of constituents



(FTSE 2014a)

Figure 4 presents the sector weights of FTSE4Good Global, FTSE4Good US and their benchmarks. The service sector includes customer service, health care and financial services, the industry sector consists of traditional industries like energy companies, utilities and material production, while the technology sector comprises technology, IT and telecommunication companies. Compared to their benchmarks, the SRI indices have larger weights in the service and technology sectors, and smaller weights in the industrial sector. For example, 33 % of the companies in FTSE4Good US are from the technology sector, compared to only 19 % in FTSE USA. The screening of FTSE4Good US has therefore resulted in 42 % more companies in the technology sector than its benchmark (FTSE USA), which is considerable.

Figure 4: Industry composition of SRI versus benchmarks



(FTSE 2014b)

These differences in average constituent size and sector composition are likely important determinants of the observed differences in return and volatility of the SRI indices and their benchmarks. For example, if the differences in sector weights were somewhat equal in the Dotcom Fall, during which technology stocks plummeted, this could help explain the inferior return of the SRI indices seen in Table 3.

7 Method

Investors do not receive higher return as compensation for taking on non-systematic risk. Therefore, the risk and return characteristics presented under the descriptive statistics does not accurately depict the financial performance of SRI. To obtain a correct estimate of the financial performance of SRI, we need to use an asset-pricing model that incorporates systematic risk factors. Consequently, we base our analysis on Fama-French's three-factor asset-pricing model (Fama and French 1992, 1993). We use this basic three-factor model to investigate SRI performance over the full sample period, as well as an expanded version with dummy variables and interaction terms to capture the distinct effects of SRI screening in periods of falling stock markets. All models are estimated using ordinary least squares (OLS).

7.1 Identification strategy

Eugene Fama and Kenneth French find that high book-to-market (value) stocks usually outperform low book-to-market (growth) stocks. They also find that small capitalization stocks tend to outperform large capitalization stocks. The rationale behind the value and size premiums is that value stocks and small capitalization stocks on average are riskier, less liquid and more prone to mispricing, but over time they tend to yield a higher return. Based on this research, they added a value factor and a size factor to the capital asset-pricing model (CAPM), resulting in model (1).

$$R_e = \alpha + \beta_1 Market + \beta_2 SMB + \beta_3 HML + \varepsilon_i \tag{1}$$

The dependent variable in model (1) is the expected excess return. The market factor is the excess return of the market over the risk-free rate. The size factor, SMB (Small minus Big), is the average return on three portfolios of small capitalization stocks minus the average return on three portfolios of large capitalization stocks. The book-to-market factor, HML (High minus Low), is the average return of two high book-to-market portfolios minus the average return of two low book-to market portfolios (French 2015). Since we are interested in how the SRI indices perform compared to their conventional benchmark indices, we need to replace the dependent variable with the differential return between the SRI indices and their benchmarks to obtain model (2). This transformation allows us to use the return of the benchmarks minus the risk-free rate (R_{BM}-r_f) as a stand-in for the market factor, but we use the regional market factors provided by Kenneth French's database (French 2015). The regional market factors are almost perfectly correlated with (R_{BM}-r_f), and the choice makes little difference.

$$R_{SRI} - R_{BM} = \alpha + \beta_1 Market + \beta_2 SMB + \beta_3 HML + \varepsilon_i \tag{2}$$

The regression output provides us with estimates for alpha (α), three betas (β) and an error term (ε). A positive (negative) alpha indicates that the SRI index has outperformed (underperformed) its respective benchmark. The three betas represent the difference in factor loadings between the SRI index and its benchmark, while the error term represents the unexplained return differences. Since the dependent variable is transformed into the return difference between an SRI index and its benchmark, we should not be surprised if the regressions have low R^2. This is because the return of an appropriate benchmark already explains much of the return variation in the SRI index.

7.2 Expanded Fama-French

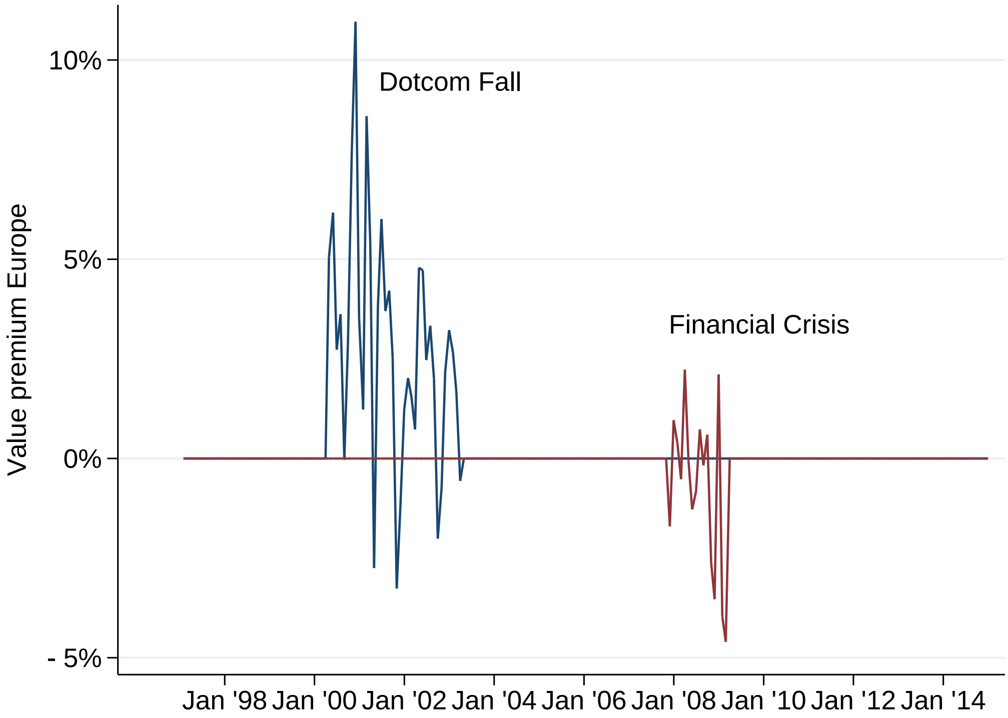
To test the performance of SRI versus conventional benchmarks in periods of falling stock markets, we can include dummy variables for the different market regimes into the model. In the Sub-period section, we divide the sample into three long periods of steady growth and two shorter periods of falling stock markets. These two periods of falling stock markets, the Dotcom Fall and the Financial Crisis, were fundamentally different in nature. Whereas the first consisted of a prolonged market correction from the tech-bubble of the late 1990s, the latter involved a global liquidity crisis and recession. Therefore, we prefer to use two separate dummy variables for these periods, while keeping the three growth periods as the base category, resulting in model (3).

$$R_{SRI} - R_{BM} = \alpha + \beta_1 Market + \beta_2 SMB + \beta_3 HML + \delta_1 D_{DotcomFall} + \delta_2 D_{FinancialCrisis} + \varepsilon_i \tag{3}$$

The addition of the dummy variables allow for different intercepts in the model, depending on whether we are looking at the Dotcom Fall, the Financial Crisis or the growth periods in the base category. However, it does not allow for different slope coefficients in the different periods. If the Fama-French factor returns vary significantly, this could lead to coefficients with a poor fit. For example, Figure 5 shows that the differences in the European value premium between the Dotcom Fall and the Financial Crisis are substantial. The average annualized European value premium is 39.01 % during the Dotcom Fall and -8.83 % during the Financial Crisis (Table A10). A difference of 47.84 %-points is large and should be adjusted for. Similar arguments can be made for the other regional Fama-French factors. To allow the coefficients to vary with the different periods, we construct interaction terms between each dummy variable and Fama-French factor, resulting in model (4).

$$\begin{aligned} R_{SRI} - R_{BM} = {} & \alpha + \beta_1 Market + \beta_2 SMB + \beta_3 HML + \delta_1 D_{DotcomFall} + \delta_2 D_{FinancialCrisis} + \\ & \gamma_1 (Market * D_{DotcomFall}) + \gamma_2 (SMB * D_{DotcomFall}) + \\ & \gamma_3 (HML * D_{DotcomFall}) + \gamma_4 (Market * D_{FinancialCrisis}) + \\ & \gamma_5 (SMB * D_{FinancialCrisis}) + \gamma_6 (HML * D_{FinancialCrisis}) + \varepsilon_i \end{aligned} \tag{4}$$

Figure 5: Monthly European value premium - Dotcom Fall and Financial Crisis



(French 2015)

The addition of the dummy variables and the interaction terms alters the interpretation of the coefficients in the model. A positive (negative) alpha now indicates that the SRI index has outperformed (underperformed) against its conventional benchmark over the three growth periods in the sample combined. On the other hand, the correct performance estimate for the Dotcom Fall and the Financial Crisis is now their respective delta coefficient (δ) plus the alpha term (α). The interpretation of the difference in the factor loadings between the SRI indices and their benchmarks has also changed. The betas now represent the growth periods, while the estimates for the Dotcom Fall and the Financial Crisis now consists of their respective gammas (γ) plus the betas (β). However, from here on out, when referring to the coefficients for the Financial Crisis or the Dotcom Fall, we will for ease of interpretation refer to ($\alpha + \delta$) and ($\beta + \gamma$) as purely alpha and beta.

7.3 Model requirements

We test whether any of the variables used in model (2) and model (4) are non-stationary, i.e. they display signs of unit roots. If the variables have unit roots, then the central limit theorem no longer applies, making large sample normal approximations invalid. This would make it impossible to trust the t-statistics and F-statistics. Furthermore, using non-stationary variables could lead to spurious regressions that cannot be trusted. For example, spurious regressions could show a causal relationship between two trending variables when there in fact is none (Wooldridge 2012). Table A3 presents the results from the Augmented Dickey-Fuller test, which rejects the null hypothesis of unit roots for all the variables, indicating that we can safely use them.

We perform postestimation tests on the residuals from each regression to ascertain whether the requirements for OLS are met. Table A4 shows the results from the Breusch-Pagan/Cook-Weisberg test for heteroskedasticity, while Table A5 presents the results from the Durbin-Watson test for first-order autocorrelation. We observe problems of both heteroskedasticity and autocorrelation in several of the regressions from model (2) and model (4). Neither heteroskedasticity nor autocorrelation leads to biased estimators, but they lead to incorrect standard errors and t-statistics. To safeguard against this, we compute Newey-West standard errors for all the regressions, which are consistent in the face of both heteroskedasticity and autocorrelation (Wooldridge 2012). We also test whether the residuals are normally distributed (Table A6). The null hypothesis of normally distributed residuals is rejected for a substantial amount of the regressions. However, we should not be too concerned about this. Since the included variables follow a stationary process and because we work with a large number of observations, we can be confident that the central limit theorem applies, which allows us to dispense with the requirement of normally distributed residuals (Wooldridge 2012).

8 Results

8.1 Main Findings

Table 4 presents the annualized differences in Fama-French factor contributions between the SRI indices and their benchmarks. The first number column presents the gross excess returns of the SRI indices over their benchmarks, while the second number column presents the annualized alphas, i.e. the factor-adjusted excess returns. We can think of these annualized alphas as the actual contribution from SRI after adjusting for systematic risk factors. The following three columns present the difference in annualized factor contributions from the market, size and value factors of the SRI indices relative to their benchmarks. Finally, the last column shows the aggregated differences in factor contributions, and whether these are positive (upwards arrow) or negative (downwards arrow).

Table 4: Fama-French Contributions

% Annualized difference returns, alphas and contributions.

		SRI-BM	Alpha	Market	Size	Value	
Full sample	FTSE4G US	-0.26	-0.37	0.42***	-0.07	-0.24**	⇑ (0.11)
(Jan'97- Dec '14)	FTSE4G UK	-0.42	-0.27	-0.23***	-0.02	0.10	⇓ (-0.15)
	FTSE4G Europe	-0.06	0.03	-0.01	-0.06***	-0.02	⇓ (-0.09)
	FTSE4G Global	0.22	0.21	0.31***	-0.03***	-0.27**	⇑ (0.01)
	KLD400	0.30	0.35	-0.02	0.08	-0.11**	⇓ (-0.05)
Financial Crisis	FTSE4G US	-4.61	-6.59**	2.97	0.90***	-1.89***	⇑ (1.98)
(Nov'07 - Feb'09)	FTSE4G UK	0.04	-1.91	1.80***	0.77***	-0.62	⇑ (1.95)
	FTSE4G Europe	-0.49	-2.65**	2.73***	0.12	-0.69	⇑ (2.16)
	FTSE4G Global	-2.82	-1.70	-0.50	0.26	-0.88**	⇓ (-1.12)
	KLD400	0.96	-3.11*	4.53***	0.55***	-1.01***	⇑ (4.07)
Dotcom Fall	FTSE4G US	-3.57	-1.55	-0.95	0.38	-1.45	⇓ (-2.02)
(April'00 - Mar'03)	FTSE4G UK	-0.80	-2.22*	0.98***	0.07**	0.37	⇑ (1.42)
	FTSE4G Europe	-2.54	-1.75	-0.16	0.10***	-0.73	⇓ (-0.79)
	FTSE4G Global	-2.69	0.59	-1.31***	-0.12	-1.85**	⇓ (-3.28)
	KLD400	-0.19	-0.28	0.36	0.38*	-0.65	⇑ (0.09)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987). The Alphas are annualized by $(1+\alpha)^{12}-1$. We derive the contributions by multiplying the annualized Fama-French factor returns (Table 5) with the regression coefficients from model (2) and (4) (Table 6). We use model (2) for the full period, while we obtain the coefficients for the Dotcom Fall and the Financial Crisis from the expanded Fama-French model (4). (SRI-BM) is calculated by summarizing the contributions from the market, size and value factors and the alphas, and will therefore deviate slightly from Table 2 and Table 3. Table inspired by Gjølberg and Johnsen (2008).

We do not find any significant alphas in either direction over the full sample period. The annualized alphas range from positive 0.35 %-points for KLD400 to negative 0.37 %-points for FTSE4Good US. This result is in alignment with previous literature, which largely concludes that there is neither a gain nor a loss from socially responsible investing. During the Financial Crisis we find that all the SRI indices underperform against their conventional benchmarks. This underperformance is significant for FTSE4Good US, FTSE4Good Europe and KLD400. The magnitude of the annualized alphas are substantial, ranging from negative 1.70 %-points for FTSE4Good Global to negative 6.59 %-points for FTSE4Good US. This same pattern is evident for the Dotcom Fall, where four out of five SRI indices underperform against their benchmarks, although only FTSE4Good UK does so significantly. These results support the underperformance hypothesis, indicating that investors suffer substantial financial losses from SRI in periods of falling markets.

The large differences between the gross excess returns and the factor-adjusted excess returns in the Dotcom Fall and the Financial Crisis, show the importance of adjusting for differences in factor contributions between the SRI indices and their benchmarks. The arrows in the last column pointing upwards (downwards) indicate that we will overestimate (underestimate) the performance of SRI if we do not adjust for these differences in factor contributions, while the numbers in parentheses indicate by how much. For example, during the Financial Crisis, KLD400 receives a positive contribution from the market factor at 4.53 %-points, a positive contribution from the size factor at 0.55 %-points and a negative contribution from the value factor at 1.01 %-points, resulting in an aggregated difference in factor contribution of 4.07 %-points. These 4.07 %-points could alternatively be obtained by investing in a conventional portfolio operating with the same factor bets as the KLD400 and does not represent an outperformance due to SRI screening. The same can of course be said in cases where the aggregated differences in factor contributions are negative, such as for FTSE4Good Global during the Dotcom Fall at negative 3.28 %-points.

For the full sample period, the aggregated differences in factor contributions are small for all the indices and display no clear trend. As a result, the gross excess returns are close to the factor-adjusted excess returns. Contrary, during the Financial Crisis four out of five SRI indices receive a large and positive aggregated factor contribution relative to their benchmarks, resulting in markedly lower factor-adjusted excess returns than gross excess returns. In other words, the factor loadings of the SRI indices proved favourable when the markets crashed. For the Dotcom Fall, there are no clear trends in the aggregated difference in factor contributions, but most of the differences are of a large magnitude. For example, FTSE4Good US can attribute 2.02 %-points of its negative gross excess return in this period to unfavourable factor returns, while the gross excess return of FTSE4Good UK would be 1.42 %-points worse after adjusting for its favourable factor returns. The following subsection dissects the aggregated difference in factor contribution for each index pair into its market, size and value components.

8.2 Differences in systematic risk factors

Table 5 shows the annualized market, size and value premiums for the full sample period, the Financial Crisis and the Dotcom Fall for all regions. Table 6 presents the differences in Fama-Fench factor loadings between the SRI indices and their benchmarks. Unless otherwise stated, all factor contributions in this subsection are annualized differences between the SRI indices and their benchmarks.

Table 5: Annualized Fama-French factors

		Market	**Size**	**Value**
Full sample	US	7.00 %	2.98 %	2.90 %
(Jan'97 - Dec'14)	Europe	6.45 %	0.48 %	5.74 %
	Global	5.88 %	0.27 %	4.83 %
Financial Crisis	US	-40.67 %	2.30 %	-11.09 %
(Nov'07 - Feb'09)	Europe	-47.69 %	-5.05%	-8.83 %
	Global	-43.24 %	-2.34 %	-5.05 %
Dotcom Fall	US	-18.19 %	5.60 %	23.72 %
(April'00 - Mar'03)	Europe	-18.60 %	-0.59 %	39.01 %
	Global	-18.94 %	1.34 %	33.27 %

Factor returns annualized by $(1+R_{Factor})^{12}-1$.
R_{Factor} is monthly factor return

Table 6: Fama-French Coefficients

Factor loadings and % annualized alphas.

		Alpha	Market	Size	Value	R^2
Full sample	FTSE4G US	-0.37	0.06***	-0.02	-0.08**	0.15
(Jan'97 - Dec'14)	FTSE4G UK	-0.27	-0.04***	-0.04	0.02	0.08
	FTSE4G Europe	0.03	0.00	-0.11***	0.00	0.20
	FTSE4G Global	0.21	0.05***	-0.12 ***	-0.06**	0.24
	KLD400	0.35	0.00	0.03	-0.04**	0.06
Financial Crisis	FTSE4G US	-6.59**	-0.07	0.31***	0.17***	0.28
(Nov'07-Feb'09)	FTSE4G UK	-1.91	-0.04***	-0.15***	0.07	0.14
	FTSE4G Europe	-2.65**	-0.06***	-0.02	0.08	0.28
	FTSE4G Global	-1.70	0.01	-0.11	0.17**	0.28
	KLD400	-3.11*	-0.11***	0.19***	0.09***	0.14
Dotcom Fall	FTSE4G US	-1.55	0.05	0.07	-0.06	0.28
(April'00-Mar'03)	FTSE4G UK	-2.22*	-0.05***	-0.12**	0.01	0.14
	FTSE4G Europe	-1.75	0.01	-0.17***	-0.02	0.28
	FTSE4G Global	0.59	0.07***	-0.09	-0.06**	0.28
	KLD400	-0.28	-0.02	0.07*	-0.03	0.14

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987). The coefficients for the Full sample period are estimated using model (2). See Table A7 for the regression output. The coefficients for the Financial Crisis and the Dotcom Fall are from model (4). See Table A8 for the regression output. Alphas for Financial Crisis and (DotcomFall) = $\alpha_{GrowthCombined} + \delta_{1(2)}$ and then annualized. Betas for Financial Crisis and (Dotcom Fall) = Coefficients for Growth Combined + $\gamma_{1/2/3/(4/5/6)}$. All $(\alpha+\delta)$ and $(\beta+\gamma)$ are tested for joint significance.

8.2.1 Value factor

The contributions from the value factor are substantial and negative for most of the SRI indices over the full sample period, as well as for the two periods of falling stock markets (Table 4). Over the full period and during the Dotcom Fall, we can explain these negative contributions by the SRI indices' tilt towards growth stocks (Table 6) combined with a positive value premium (Table 5). During the Dotcom Fall, the negative contributions from the value factor are larger than for the full period, which is because the value premium soared in this period. The large value premium can be explained by the dramatic fall in the valuation of typical growth sectors, such as IT, telecom and technology, in the aftermath of the tech-bubble in the late 1990s.

During the Financial Crisis, we observe that the SRI indices shift their tilt towards value stocks. We do not have access to constituent data for the SRI indices and their benchmarks, so a discussion of the reason behind this shift would be purely speculative. Since value stocks underperform growth stocks in this period, the SRI indices unanimously receive negative contributions from the value factor. For example, FTSE4Good US can attribute 1.89 %-points, significant at the 1 %-level, of its negative gross excess return to the unfavorable loading on the value factor in this period.

8.2.2 Size factor

The SRI indices display a tilt towards large capitalization stocks for the full period. The exception is KLD400, which has a larger, but insignificant, weighting in small capitalization stocks than its benchmark. This is likely due to the specific diversification requirement of KLD400, discussed in section 5.1. The resulting contributions from the size factor are small for all the indices. For example, the tilt towards large capitalization stocks of FTSE4Good Global relative to its benchmark is significant at the 1 % level, but resulted in a negative contribution from the size factor of a mere 0.03 %-points.

In the Financial Crisis, we observe that all the SRI indices receive a positive contribution from the size factor, significant for FTSE4Good US, FTSE4Good UK and KLD400. The positive contributions of the European, U.K. and Global SRI indices can be explained by their tilt toward large companies, which outperformed small companies in Europe and globally. Contrary, large companies underperform small companies in the U.S. in this period. For KLD400, which consistently tilt towards small companies throughout the sample, this yields a positive factor contribution of 0.55 %-points. More surprisingly, we find that FTSE4Good US changed its tilt significantly towards small companies, resulting in a positive contribution from the size factor at 0.90 %-points. This shift towards small companies can partially be explained by the corresponding shift towards value companies. The correlation between the value and size premium during the Financial Crisis is positive at 34.98 %, illustrating that value companies are typically smaller than growth companies. Therefore, when FTSE4Good US changed its tilt towards value stocks, it is reasonable that the average size of its constituents should fall as well.

During the Dotcom Fall, we observe that small companies outperform large companies in the U.S. and globally, but not in Europe. This inferior return of small companies in Europe is consistent with Lu and Chollete (2010) who shows that the positive factor contribution from loading on small companies holds over the long run, but may have shorter-run regime shifts. Combined with the tilts of FTSE4Good UK and FTSE4Good Europe towards large companies during the Dotcom Fall, the resulting contributions from the size factor are positive and significant for these indices. The tilts of the remaining SRI indices remain mostly the same as for the full period.

8.2.3 Market factor

Most of the contributions from the market factor are of a larger magnitude than those from the size and value factors. The main reason for these large contributions are the substantial factor returns from loading on the market factor and not the loading differences themselves. For example, the coefficient on the market factor and the coefficient on the size factor for FTSE4Good UK over the full sample period is equal at -0.04. However, the average return from loading on the market factor at 6.45 % p.a. is considerably larger than the average return from loading on the size factor at 0.48 % p.a.. As a result, the negative contribution from the market factor at 0.23 %-points is noticeably larger than the negative contribution from the size factor at 0.02 %-points.

For the full sample period and the Dotcom Fall, there are no clear trends in the loadings on the market factor. In the Financial Crisis however, four out of five SRI indices were less exposed to the market than their benchmarks, three of them significantly, which proved advantageous when the markets crashed. For example, KLD400 receive a positive factor contribution of 4.53 %-points due to this lower market exposure, significant at the 1 %-level. The lower market exposure in the Financial Crisis can partially be explained by the tilts on the size factor and the value factor in this period. Fama and French (1993) show that growth stocks generally have a higher market beta than value stocks, while large stocks generally have a lower market beta than small stocks. The SRI indices display a tilt towards value stocks during the Financial Crisis, which can explain some of the reduced market exposure. In the same fashion, the tilt towards large stocks of the SRI indices outside of the U.S. may have further reduced their market exposure, while the U.S. indices' tilt towards small stocks may have increased theirs.

9 Discussion

The results imply that there are no financial gains nor losses from SRI if the investment horizon is long, indicating that the benefits of screening for ESG criteria offsets the disadvantages. However, the results also point toward substantial financial losses from SRI in the Financial Crisis and in the Dotcom Fall. This means that the adverse effects of SRI, such as the increased idiosyncratic risk from the reduced investment universe and skewed sector tilts, outweighs the benefits, resulting in SRI underperforming conventional investments in periods of falling stock markets.

9.1 Practical implications

The above findings have important implications for the types of investors that should be involved in SRI. For example, most pension funds and university endowment funds have an infinite investment horizon and are not as preoccupied with the short-term fluctuations of their portfolios. If the concept of SRI sounds appealing to the principals of these funds, then they can invest without having to fear inferior financial returns. Contrary, investors with a shorter investment horizon, investors with lower risk tolerance and institutional investors under heavy public scrutiny should all be wary of SRI. These types of investors might find it difficult enough to weather periods of falling markets without adding the idiosyncratic risk caused by SRI screening.

9.2 Theoretical implications

Our findings expand on the previous literature on SRI performance by adjusting for both systematic risk factors and different market regimes simultaneously. When only adjusting for systematic risk factors, Kurtz and DiBartolomeo (2011) find that the differences in performance between KLD400 and its benchmark are purely factor driven. Likewise, when adjusting for different market regimes, but not adjusting for systematic risk factors, Managi et al. (2012) find no under- or outperformance by the SRI indices they investigate. Therefore, the underperformance of the SRI indices in periods of falling markets revealed in our results indicate that adjusting for both systematic risk factors and different market regimes together might be beneficial.

9.3 Limitations and further research

A major problem when working with this study was the lack of holding data for the indices. These data are tremendously expensive and our attempts at obtaining these free of charge from FTSE and MSCI have been fruitless. Future researchers should attempt to obtain holding data for all the indices to be able to control for industries in the regression models. Furthermore, with the increased popularity of SRI follows an increased number of SRI indices. If the number of trustworthy SRI indices reach a certain level, one should consider using panel data techniques. Specifically, one would then be able to combine the effects of the different screening systems between each index family into one aggregate. This panel data regression would yield a single alpha term, representing a more accurate measurement of SRI performance. Lastly, we have not attempted to discern whether performance varies between SRI indices exclusively using positive or negative screening. This is because all the trustworthy SRI indices with a comparable benchmark we could find utilize both. Hopefully, future research will be able to overcome this problem.

10 Conclusion

In this study, we expand on previous research on SRI performance by investigating the return difference of five regionally diverse sets of SRI indices and their benchmarks over an 18-year period from 1997 to 2014. Through a Fama-French three-factor model, we account for differences in systematic risk factors between the SRI indices and their benchmarks, specifically the tendency of the SRI indices to tilt towards large growth stocks. The results show that there are no financial gains nor losses from investing in the SRI indices instead of their benchmarks if the investment horizon is long, indicating that the benefits of SRI screening offsets the disadvantages.

Among these disadvantages is an increase in the idiosyncratic risk caused by the reduced investment universe and skewed sector tilts from SRI screening. We hypothesize that this increased idiosyncratic risk is more likely to translate into inferior risk-adjusted returns in periods of falling markets, when having a well-diversified portfolio is crucial. To test this hypothesis, we expand the Fama-French three-factor model with dummy variables for the Dotcom Fall in the early 2000s and the Financial Crisis of 2007 to 2009. Furthermore, to adjust for the large variations in the market, size and value premiums in these periods, we include interaction terms in the model. The results show large financial losses from investing in the SRI indices instead of their benchmarks in both periods of falling markets.

The findings in this study implies that SRI performs on par with conventional investments in the long run and therefore might be a suitable investment option for investors with an infinite investment horizon, such as pension funds and university endowment funds. However, the inferior risk-adjusted returns in periods of falling markets should make investors concerned with short-term portfolio fluctuations wary of SRI.

A limitation in the study is the lack of holding data, preventing us from controlling for differences in industry weights between the SRI indices and their benchmarks. Another limitation is that all the SRI indices in our sample use both positive and negative screening. By using the methods presented in this study, and by including indices that only use positive or negative screening, future researchers can attempt to discern how the different screening methods affect SRI performance in the long run and in times of falling markets.

References

Bank Of England. 2015. "Bank of England's Statistical Interactive Database." Visited on 03/11/2015. http://www.bankofengland.co.uk/boeapps/iadb/Index.asp?first=yes&SectionRequired=I&HideNums=-1&ExtraInfo=true&Travel=NIx.

Belghitar, Yacine, Ephraim Clark, and Nitin Deshmukh. 2014. "Does it pay to be ethical? Evidence from the FTSE4Good." *Journal of Banking & Finance* 47:54–62.

Bloomberg, L.P. *Bloomberg Terminal NHH*.

Boon, Junyan, Jacquelyn E Humphrey, and Geoffrey J Warren. 2013. "Skill and style in socially responsible portfolios: A holdings-based analysis."

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

Commission, EU. 2011. *A renewed EU strategy 2011-2014 for Corporate Social Responsibility*. Visited on 15/04/2015. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0681:FIN:EN:PDF.

DataStream. 2015. *Thomson Reuter's DataStream*.

Derwall, Jeroen, et al. 2005. "The eco-efficiency premium puzzle." *Financial Analysts Journal* 61 (2): 51–63.

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

Elliot, Larry. 2012. "Three myths that sustain the economic crisis." Visited on 08/01/2015. http://www.theguardian.com/business/economics-blog/2012/aug/05/economic-crisis-myths-sustain.

EUROSIF. 2015. *European SRI study 2014*. Visited on 01/02/2015. http://www.eurosif.org/publication/view/european-sri-study-2014/.

Fama, Eugene F, and Kenneth R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of financial economics* 33 (1): 3–56.

— . 2012. "Size, value, and momentum in international stock returns." *Journal of financial economics* 105 (3): 457–472.

— . 1992. "The cross-section of expected stock returns." *the Journal of Finance* 47 (2): 427–465.

French, Kenneth R. 2015. "Data Library." Visited on 02/01/2015. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Friedman, Milton. 1970. "The social responsibility of business is to increase its profits." *New York Times Magazine* 13:32–33.

FTSE. 2014a. "FTSE4 Good Global Index Fact Sheet." Visited on 01/04/2015. http://www.ftse.com/Analytics/FactSheets/temp/0aa4d0ae-73d3-4fe1-a0f1-066ed5df99bc.pdf.

— . 2014b. "FTSE4Good Index Series Fact Sheet." Visited on 01/04/2015. http://www.ftse.com/Analytics/FactSheets/temp/68ecb123-1cdb-4624-93c4-92a687964946.pdf.

— . 2014c. "Integrating ESG into investments and stewardship." Visited on 01/04/2015. http://www.ftse.com/products/downloads/FTSE-ESG-Methodology-and-Usage-Summary-Short.pdf?825.

Gjølberg, Ole, and Thore Johnsen. 2008. "Etisk forvaltning av Statens Pensjonsfond Utland: En oppdatert analyse." *On behalf of the Norwegian Ministry of Finance*.

— . 2003. "Evaluering av etisk forvaltning: Metode, resultater og kostnader." *Attachment 11 in NOU 2003:22*.

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

Humphrey, Jacquelyn E, and David T Tan. 2014. "Does it Really Hurt to be Responsible?" *Journal of business ethics* 122 (3): 375–386.

Humphrey, Jacquelyn, and David Tan. 2011. "The Many Faces of Socially Responsible Investing-Does the Screening Mechanism Affect the Risk and Return of Mutual Funds?" *Available at SSRN 1913987*.

Israelsen, Craig L. 2003. "Sharpening the Sharpe ratio." *Financial Planning* 33 (1): 49–51.

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

Kurtz, Lloyd, and Dan DiBartolomeo. 2011. "The long-term performance of a social investment universe." *The Journal of Investing* 20 (3): 95–102.

Lu, Ching-Chih, and Loran Chollete. 2010. *Bankruptcy and the size effect*. Tech. rep. University of Stavanger.

Managi, Shunsuke, Tatsuyoshi Okimoto, and Akimi Matsuda. 2012. "Do socially responsible investment indexes outperform conventional indexes?" *Applied Financial Economics* 22 (18): 1511–1527.

Margolis, Joshua D, Hillary Anger Elfenbein, and James P Walsh. 2007. "Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance." *Ann Arbor* 1001:48109–1234.

Mortier, Ophelie. 2014. "The added value of integrating extra-financial criteria in an investment process." Visited on 05/29/2015. https://yoursri.com/media-new/download/2014_11_topicofthemonth_extrafinancialcriteriainvestmentprocess-2.pdf.

MSCI. 2014. "MSCI KLD400 Social Index Fact Sheet per December 2014." Visited on 15/02/2015. https://www.msci.com/resources/factsheets/index_fact_sheet/msci-kld-400-social-index.pdf.

Nestlé. 2011a. Visited on 30/05/2015. http://www.nestle.com/aboutus/ask-nestle/answers/nestle-boycott.

— . 2011b. Visited on 30/05/2015. http://www.nestle.com/aboutus/ask-nestle/answers/what-is-ftse4good.

Newey, Whitney K, and Kenneth D West. 1987. "Hypothesis testing with efficient method of moments estimation." *International Economic Review*: 777–787.

Ng, Serena, and Pierre Perron. 2001. "Lag length selection and the construction of unit root tests with good size and power." *Econometrica* 69 (6): 1519–1554.

RBC, Global Asset Management. 2012. "Does Socially Responsible Investing Hurt Investment Returns?" Visited on 01/03/2015. http://funds.rbcgam.com/_assets-custom/pdf/RBC-GAM-does-SRI-hurt-investment-returns.pdf.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang. 2007. "Socially responsible investments: Methodology, risk exposure and performance."

— . 2008. "The price of ethics and stakeholder governance: The performance of socially responsible mutual funds." *Journal of Corporate Finance* 14 (3): 302–322.

Schröder, Michael. 2007. "Is there a difference? The performance characteristics of SRI equity indices." *Journal of Business Finance & Accounting* 34 (1-2): 331–348.

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

USSIF. 2014. "Report on US Sustainable, Responsible and Impact Investing Trends." Visited on 01/10/2015. http://www.ussif.org/trends.

Wooldridge, Jeffrey. 2012. *Introductory econometrics: A modern approach*. Cengage Learning.

A Appendix

A.1 Descriptive statistics

Table A1: Sharpe Ratios

	Full Sample	**Tech Bubble**	**Dotcom Fall**	**Growth**	**Financial Crisis**	**Recovery/Growth**
	Jan'97-Dec'14	Jan'97-Mar '00	April'00 - Mar'03	April'03-Oct'07	Nov'07-Feb'09	Mar'09-Dec'14
FTSE4G US	0.39	1.48	-1.10	1.34	-2.16	1.76
FTSE USA	0.44	1.38	-1.07	1.71	-2.18	1.71
FTSE4G UK	0.27	1.00	-1.26	1.69	-1.85	1.19
FTSE All-Share	0.29	0.88	-1.17	1.92	-1.86	1.18
FTSE4G Europe	0.37	1.76	-1.31	2.03	-2.52	1.32
FTSE Dev Europe	0.39	1.62	-1.28	2.22	-2.47	1.29
FTSE4G Global	0.35	1.45	-1.23	2.09	-2.14	1.23
FTSE Dev World	0.36	1.12	-1.23	2.31	-2.11	1.29
KLD400	0.44	1.53	-1.09	1.38	-2.20	1.70
MSCI USA	0.43	1.39	-1.12	1.66	-2.18	1.70

Sharpe ratio = (μ_p - r_f)/ σ_p

Table A2: Descriptive statistics

% Annualized Mean return and Standard deviation

	Full sample		Tech Bubble		Dotcom Fall		Growth		Financial Crisis		Recovery/growth	
	(Jan'97	- Dec'14)	(Jan'97	- Mar'00)	(Arpil'00	- Mar'03)	(April'03	- Oct'07)	(Nov'07	- Feb'09)	(Mar'09	- Dec'14)
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
FTSE4G US	9.02 %	17.12%	33.62 %	19.31 %	-18.15 %	19.79 %	14.81 %	8.89 %	-42.39 %	20.49 %	24.40 %	13.84 %
FTSE USA	9.31 %	15.66 %	28.51 %	17.10 %	-15.14 %	17.51 %	16.89 %	8.18 %	-39.71 %	19.09 %	22.93 %	13.42%
Diff	-0.29%	1.46 %	5.11 %	2.21 %	-3.01 %	2.28 %	-2.08 %	0.71 %	-2.68 %	1.40 %	1.47 %	0.42 %
FTSE4G UK	7.20%	13.98 %	19.71 %	13.60 %	-14.88 %	15.50 %	18.07 %	8.00 %	-30.89 %	18.81 %	15.86 %	13.01 %
FTSE All-Share	7.63%	14.30 %	18.64 %	14.26 %	-14.21 %	16.04 %	20.42 %	8.26 %	-31.43 %	19.09 %	15.82 %	13.09 %
Diff	-0.43 %	-0.32 %	1.07 %	-0.66 %	-0.67 %	-0.54 %	-2.35 %	-0.26 %	0.54 %	-0.28 %	0.04 %	-0.08 %
FTSE4G Europe	8.34 %	16.80%	37.30 %	19.34 %	-21.38 %	19.30 %	21.13 %	9.23 %	-42.17 %	17.93 %	17.49 %	13.18 %
FTSE Dev Europe	8.38%	16.32%	33.00 %	18.28 %	-19.42 %	18.24 %	22.82 %	9.20 %	-42.51 %	18.36 %	17.13 %	13.22 %
Diff	-0.04%	0.48%	4.30 %	1.06 %	-1.96 %	1.06 %	-1.69 %	0.03 %	0.34 %	-0.43 %	0.36 %	-0.04 %
FTSE4G Global	8.33%	17.02%	29.10 %	16.70 %	-18.38 %	17.87 %	21.64 %	8.96 %	-44.13 %	21.51 %	19.82 %	16.09 %
FTSE Dev World	8.10%	15.87%	22.16 %	15.36 %	-16.20 %	16.07 %	22.73 %	8.57 %	-42.34 %	21.00 %	19.52 %	15.09 %
Diff	0.23%	1.15%	6.94 %	1.34 %	-2.18 %	1.80 %	-1.09 %	0.39 %	-1.79 %	0.51 %	0.30 %	1.00 %
KLD400	9.48%	16.03%	32.90 %	18.24 %	-16.01 %	17.90 %	15.17 %	8.85 %	-38.72 %	18.48 %	22.84 %	13.41 %
MSCI USA	9.16%	15.71%	28.99 %	17.24 %	-15.92 %	17.33 %	16.72 %	8.32 %	-39.90 %	19.16 %	22.95 %	13.44 %
Diff	0.32%	0.31%	3.91 %	1.00 %	-0.09 %	0.57 %	-1.55 %	0.53 %	1.18 %	-0.68 %	-0.11 %	-0.03 %

Mean return is annualized by $(1+R_{Month})^{12}-1$. R_{Month} is the arithmetic monthly mean.
Stdev is annualized by $StDev_{Month} * \sqrt{12}$. $StDev_{Month}$ is the arithmetic monthly mean.

A.2 Diagnostic tests

Table A3: Augmented Dickey-Fuller test for unit roots

Variable	Number[1] of lags	Test[2] statistic	Unit[3] roots
Diff FTSE4G US	11	-3.381	Rejected***
Diff FTSE4G UK	13	-2.396	Rejected**
Diff FTSE4G Europe	11	-2.252	Rejected**
Diff FTSE4G Global	11	-1.892	Rejected*
Diff KLD400	13	-3.039	Rejected***
Market U.S.	11	-3.284	Rejected***
Size U.S.	13	-3.938	Rejected***
Value U.S.	10	-3.697	Rejected***
Market Europe	14	-3.388	Rejected***
Size Europe	12	-3.454	Rejected***
Value Europe	10	-3.150	Rejected***
Market Global	14	-3.447	Rejected***
Size Global	14	-2.587	Rejected**
Value Global	10	-3.270	Rejected***

[1] We find the optimal number of lags by running a Dickey-Fuller Generalized Least Squares (DF-GLS) test on each variable and using the Modified Akaike's Information Criterion (MAIC) (Ng and Perron 2001).
[2] Obtained from the ADF test with the MAIC optimal number of lags. None of the variables have a significant trend or constant term, so we run ADF without adding trend and by removing the constant term.
[3]Test statistics are compared to the Dickey-Fuller distribution. H_0: Unit roots
* = significant 10 %, ** = significant 5%, ***= significant 1%

Table A4: Breusch-Pagan / Cook-Weisberg
Test for heteroskedasticity

	FF model (2)		Expanded FF model (4)	
	P-value	**Homoskedasticity**	**P-value**	**Homoskedasticity**
FTSE4G US	0.9614	Not Rejected	0.4680	Not Rejected
FTSE4G UK	0.5690	Not Rejected	0.1466	Not Rejected
FTSE4G Europe	0.0049	Rejected***	0.0724	Rejected*
FTSE4G Global	0.4380	Not Rejected	0.2665	Not Rejected
KLD400	0.2274	Not Rejected	0.7994	Not Rejected

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
H_0: Constant variance

Table A5: Durbin-Watson
Test for first-order autocorrelation

	FF model (2)			Expanded FF model (4)		
	d	**Positive**[1]	**Negative**[2]	**d**	**Positive**[1]	**Negative**[2]
FTSE4G US	2.03	Not Rejected	Not Rejected	2.08	Not Rejected	Not Rejected
FTSE4G UK	2.63	Not Rejected	Rejected	2.63	Not Rejected	Rejected
FTSE4G Europe	2.26	Not Rejected	Inconclusive	2.36	Not Rejected	Rejected
FTSE4G Global	1.95	Not Rejected	Not Rejected	1.95	Not Rejected	Not Rejected
KLD400	1.92	Not Rejected	Not Rejected	1.98	Not Rejected	Not Rejected

[1] H_0 = No positive first-order autocorrelation
[2] H_0 = No negative first-order autocorrelation
"d" is the Durbin-Watson test statistic. We use (4-d) to test for first-order negative autocorrelation.
Model (2): k=4 , n=216, α=0.05 gives d_{lower}= 1.73 and d_{upper}=1.81
Model (4): k=12 , n=216, α=0.05 gives d_{lower}= 1.67 and d_{upper}=1.87

Table A6: Shapiro-Wilk normality test

	FF model (2)	**Expanded FF model (4)**
FTSE4G US	Not Rejected	Not Rejected
FTSE4G UK	Rejected***	Rejected***
FTSE4G Europe	Rejected***	Rejected***
FTSE4G Global	Rejected**	Rejected***
KLD400	Not Rejected	Not Rejected

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
H_0: Normally distributed residuals

A.3 Fama-French

Table A7 and Table A8 presents the output from regression model (2) and (4), respectively. The outputs are used to construct the coefficients and annualized alphas in Table A9. To get the contributions in Table A11, the coefficients are multiplied with the annualized Fama-French factors in Table A10.

Table A7: Output Model (2)

Factor loadings and % monthly alphas. T-values in parentheses.

Full sample	α	β_{Market}	β_{Size}	β_{Value}	R^2
FTSE4G US	-0.031	0.06***	-0.02	-0.08**	0.15
	(-0.49)	(3.48)	(-0.89)	(-2.44)	
FTSE4G UK	-0.023	-0.04***	-0.04	0.02	0.08
	(-0.60)	(-4.21)	(-1.10)	(1.27)	
FTSE4G Europe	0.003	0.00	-0.11***	0.00	0.20
	(0.08)	(-0.11)	(-5.27)	(-0.29)	
FTSE4G Global	0.018	0.05***	-0.12 ***	-0.06**	0.24
	(0.37)	(5.03)	(-3.73)	(-2.04)	
KLD400	0.029	0.00	0.03	-0.04**	0.06
	(0.65)	(-0.24)	(1.63)	(-2.15)	

* p < 0.10 , ** p < 0.05 , *** p < 0.01
All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987).
Alphas and deltas (δ) are monthly percentages.

Table A8: Output Model (4)

Betas (β), interaction terms (γ) and % monthly alphas (α) and deltas (δ) from model (4).
T-values in parentheses.

Growth Combined	α	β_{Market}	β_{Size}	β_{Value}	R^2
FTSE4G US	0.003	0.06***	-0.08***	-0.14***	0.28
	(0.05)	(2.67)	(-2.99)	(-3.62)	
FTSE4G UK	0.029	-0.04***	-0.01	0.03*	0.14
	(0.59)	(-3.76)	(-0.15)	(1.70)	
FTSE4G Europe	0.065	-0.01	-0.11***	0.04**	0.28
	(1.46)	(-1.20)	(-4.01)	(2.18)	
FTSE4G Global	0.023	0.05***	-0.13***	-0.07*	0.28
	(0.38)	(3.42)	(-4.12)	(-1.67)	
KLD400	-0.004	0.02	-0.00	-0.07***	0.14
	(-0.09)	(1.11)	(-0.08)	(-2.65)	
Financial Crisis	δ_1	γ_1	γ_2	γ_3	R^2
FTSE4G US	-0.569**	-0.13***	0.39***	0.31***	0.28
	(-1.99)	(-1.67)	(4.24)	(3.90)	
FTSE4G UK	-0.190	0.00	-0.14**	0.04	0.14
	(-1.64)	(0.27)	(-2.09)	(0.57)	
FTSE4G Europe	-0.288**	-0.05**	0.09	0.04	0.28
	(-2.44)	(-2.04)	(1.40)	(0.66)	
FTSE4G Global	-0.166	-0.04	0.02	0.24**	0.28
	(-1.04)	(-1.50)	(0.18)	(2.48)	
KLD400	-0.259	-0.13***	0.19***	0.16***	0.14
	(-1.57)	(-3.47)	(4.35)	(4.24)	
Dotcom Fall	δ_2	γ_4	γ_5	γ_6	R^2
FTSE4G US	-0.133	-0.01	0.15***	0.08	0.28
	(-0.71)	(-0.10)	(3.02)	(1.35)	
FTSE4G UK	-0.216*	-0.01	-0.11	-0.02	0.14
	(-1.76)	(-0.48)	(-1.57)	(-0.90)	
FTSE4G Europe	-0.212**	0.02	-0.06	-0.06**	0.28
	(-1.90)	(1.30)	(-1.52)	(-2.06)	
FTSE4G Global	0.026	0.02	0.04	0.01	0.28
	(0.19)	(0.56)	(0.62)	(0.32)	
KLD400	-0.019	-0.04	0.07	0.04	0.14
	(-0.13)	(-1.00)	(1.47)	(0.93)	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987).
Alphas and deltas (δ) are monthly percentages.

Table A9: Fama-French Coefficients including Growth Combined

Factor loadings and % annualized alphas.

		Alpha	Market	Size	Value	R^2
Full Sample	FTSE4G US	-0.37	0.06***	-0.02	-0.08**	0.15
(Jan'97 - Dec'14)	FTSE4G UK	-0.27	-0.04***	-0.04	0.02	0.08
	FTSE4G Europe	0.03	0.00	-0.11***	0.00	0.20
	FTSE4G Global	0.21	0.05***	-0.12 ***	-0.06**	0.24
	KLD400	0.35	0.00	0.03	-0.04**	0.06
Financial Crisis	FTSE4G US	-6.59**	-0.07	0.31***	0.17***	0.28
(Nov'07-Feb'09)	FTSE4G UK	-1.91	-0.04***	-0.15***	0.07	0.14
	FTSE4G Europe	-2.65**	-0.06***	-0.02	0.08	0.28
	FTSE4G Global	-1.70	0.01	-0.11	0.17**	0.28
	KLD400	-3.11*	-0.11***	0.19***	0.09***	0.14
Dotcom Fall	FTSE4G US	-1.55	0.05	0.07	-0.06	0.28
(April'00-Mar'03)	FTSE4G UK	-2.22*	-0.05***	-0.12**	0.01	0.14
	FTSE4G Europe	-1.75	0.01	-0.17***	-0.02	0.28
	FTSE4G Global	0.59	0.07***	-0.09	-0.06**	0.28
	KLD400	-0.28	-0.02	0.07*	-0.03	0.14
Growth Combined	FTSE4G US	0.03	0.06***	-0.08***	-0.14***	0.28
	FTSE4G UK	0.35	-0.04***	-0.01	0.03*	0.14
	FTSE4G Europe	0.78	-0.01	-0.11***	0.04**	0.28
	FTSE4G Global	0.28	0.05***	-0.13***	-0.07*	0.28
	KLD400	-0.05	0.02	0.00	-0.07***	0.14

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987).

The coefficients for the Full sample period are estimated using model (2). See Table A7 for the regression output.

The coefficients for the Financial Crisis and the Dotcom Fall are from model (4). See Table A8 for the regression output

Alphas for Financial Crisis and (DotcomFall) = $\alpha_{GrowthCombined} + \delta_{1(2)}$ and then annualized.

Betas for Financial Crisis and (Dotcom Fall) = Coefficients for Growth Combined + $\gamma_{1/2/3/(4/5/6)}$.

All $(\alpha + \delta)$ and $(\beta + \gamma)$ are tested for joint significance.

Table A10: Annualized Fama-French factors

		Market	Size	Value
Full sample	US	7.00 %	2.98 %	2.90 %
(Jan'97 - Dec'14)	Europe	6.45 %	0.48 %	5.74 %
	Global	5.88 %	0.27 %	4.83 %
Tech Bubble	US	21.99 %	-0.39 %	-10.25 %
(Jan'97 - Mar'00)	Europe	15.54 %	-2.98 %	-5.29 %
	Global	15.90 %	-4.98 %	-8.19 %
Dotcom Fall	US	-18.19 %	5.60 %	23.72 %
(April'00 - Mar'03)	Europe	-18.60 %	-0.59 %	39.01 %
	Global	-18.94 %	1.34 %	33.27 %
Growth	US	14.30 %	3.98 %	5.75 %
(April'03- Oct'07)	Europe	28.77 %	4.79 %	7.98 %
	Global	20.67 %	4.38 %	6.07 %
Financial Crisis	US	-40.67 %	2.30 %	-11.09 %
(Nov'07 - Feb'09)	Europe	-47.69 %	-5.05%	-8.83 %
	Global	-43.24 %	-2.34 %	-5.05 %
Recovery	US	23.30 %	2.93 %	2.08%
(Mar'09-Dec'14)	Europe	17.33 %	0.98 %	-0.85 %
	Global	19.41 %	0.17 %	0.92 %
Growth Combined	US	19.91 %	2.48 %	0.20 %
	Europe	20.62 %	1.2 8%	0.94 %
	Global	18.98 %	0.30 %	0.35 %

Factor returns annualized by $(1+R_{Factor})^{12}-1$.
R_{Factor} is monthly factor return

Table A11: Fama-French Contributions including, Growth Combined

		SRI-BM	Alpha	Market	Size	Value
Full sample	FTSE4G US	-0.26	-0.37	0.42***	-0.07	-0.24**
(Jan'97- Dec '14)	FTSE4G UK	-0.42	-0.27	-0.23***	-0.02	0.10
	FTSE4G Europe	-0.06	0.03	-0.01	-0.06***	-0.02
	FTSE4G Global	0.22	0.21	0.31***	-0.03***	-0.27**
	KLD400	0.30	0.35	-0.02	0.08	-0.11**
Financial Crisis	FTSE4G US	-4.61	-6.59**	2.97	0.90***	-1.89***
(Nov'07 - Feb'09)	FTSE4G UK	0.04	-1.91	1.80***	0.77***	-0.62
	FTSE4G Europe	-0.49	-2.65**	2.73***	0.12	-0.69
	FTSE4G Global	-2.82	-1.70	-0.50	0.26	-0.88**
	KLD400	0.96	-3.11*	4.53***	0.55***	-1.01***
Dotcom Fall	FTSE4G US	-3.57	-1.55	-0.95	0.38	-1.45
(April'00 - Mar'03)	FTSE4G UK	-0.80	-2.22*	0.98***	0.07**	0.37
	FTSE4G Europe	-2.54	-1.75	-0.16	0.10***	-0.73
	FTSE4G Global	-2.69	0.59	-1.31***	-0.12	-1.85**
	KLD400	-0.19	-0.28	0.36	0.38*	-0.65
Growth Combined	FTSE4G US	0.94	0.03	1.13***	-0.19***	-0.03***
	FTSE4G UK	-0.51	0.35	-0.88***	-0.01	0.03*
	FTSE4G Europe	0.43	0.78	-0.26	-0.13***	0.04**
	FTSE4G Global	1.23	0.28	1.01***	-0.04***	-0.02*
	KLD400	0.27	-0.05	0.33	0.00	-0.01***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All regressions use Newey-West standard errors with 5 lags, where the number of lags is determined by $4(n/100)^{2/9}$ (Newey and West 1987).
The annulized alphas are retrieved from Table A9
We derive the contributions by multiplying the annualized Fama-French factor returns (Table A10) with the regression coefficients from model (2) and (4) (Table A9). We use model (2) for the full period, while we obtain the coefficients for the Dotcom Fall and the Financial Crisis from the expanded Fama-French model (4).