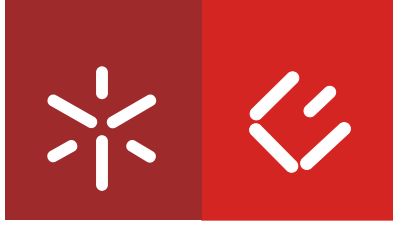**Universidade do Minho**
Escola de Economia e Gestão

Patrícia Araújo Pereira

**The impact of socially responsible investing
on the performance of European bond portfolios**

abril de 2018

Universidade do Minho
Escola de Economia e Gestão

Patrícia Araújo Pereira

The impact of socially responsible investing on the performance of European bond portfolios

Dissertação de Mestrado
Mestrado em Finanças

Trabalho efetuado sob a orientação da
Professora Doutora Maria do Céu Cortez

abril de 2018

DECLARAÇÃO

Nome: Patrícia Araújo Pereira

Endereço eletrónico: patriciaar.pereira@gmail.com

Número do Cartão de Cidadão: 13429852

Título da dissertação: The impact of socially responsible investing on the performance of European bond portfolios

Orientador: Professora Doutora Maria do Céu Cortez

Ano de Conclusão: 2018

Designação do Mestrado: Finanças

Escola: Escola de Economia e Gestão

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, 30 de Abril de 2018

Assinatura: _____

ACKNOWLEDGEMENTS

I am very proud for having this moment to acknowledge the people who contribute to the success of this dissertation.

I would especially like to thank Professor Maria do Céu Cortez, my supervisor, for sharing information and knowledge and being patient, but mainly for asking the right questions. These insightful questions allowed me to do a deep research and reach important and interesting conclusions. I am thankful for the guidance. In addition, I would also like to thank her for being available to help me share the results of this dissertation in conferences as well as to Professor Gilberto Loureiro for supporting it as the Director of Master in Finance.

I would also like to acknowledge Professor Florinda Silva for her advices on the research seminars and EEG Research Day. Furthermore, I would like to thank the iBoxx team for sharing the information I needed.

Clearly, I am grateful to my parents and friends, especially to my sister Filipa for looking after our grandmother when I needed further time to finish this journey. Thank you Filipa for being by my side.

ABSTRACT

This dissertation investigates the performance of socially screened bond portfolios of 189 Eurozone companies between 2003 and 2016. Bond portfolios are formed on the basis of an aggregate measure of corporate social responsibility (CSR) as well as on specific dimensions of CSR: the environment, social and governance dimensions. The high- and low-rated portfolios consist of the best and worst socially rated companies with bonds in each period, respectively. Our results suggest that the performance of high-rated bonds is not statistically different from that of low-rated bonds. We further analyze the evolution of bond portfolio performance over time. The results indicate that in an earlier stage portfolios of high-rated bonds outperformed portfolios of low-rated bonds. Yet, over time this outperformance diminishes and loses statistical significance. These results suggest that the errors-in-expectations hypothesis and the shunned-stock hypothesis are not only useful to explain the performance of equity portfolios but they are also useful in explaining the performance of fixed-income securities over time. When analyzing the performance of SRI bond portfolios in different market states, the results show no performance differences in periods of recessions compared to expansions.

RESUMO

Esta dissertação investiga o desempenho de carteiras de obrigações socialmente responsáveis de 189 empresas da zona euro entre 2003 e 2016. As carteiras de obrigações são formadas com base numa medida agregada de responsabilidade social empresarial bem como em dimensões específicas de responsabilidade social empresarial: dimensão ambiental, social e de governação. As carteiras com classificações altas e baixas incluem as melhores e piores empresas (em termos de responsabilidade social) com obrigações em cada período, respetivamente. Os nossos resultados sugerem que o desempenho de obrigações com classificações altas não é estatisticamente diferente do das obrigações com classificações baixas. Analisamos, adicionalmente, a evolução do desempenho das carteiras de obrigações ao longo do tempo. Os resultados indicam que numa primeira fase as carteiras com classificações altas superaram o desempenho das carteiras com avaliações baixas. No entanto, ao longo do tempo o seu desempenho diminuiu e perdeu significância estatística. Estes resultados sugerem que a errors-in-expectations hypothesis e shunned-stock hypothesis não são apenas úteis para explicar o desempenho de carteiras de ações mas são também úteis para explicar o desempenho de carteiras de obrigações ao longo do tempo. Ao analisar o desempenho de carteiras de obrigações socialmente responsáveis em diferentes estados do mercado, os resultados não mostram diferenças no desempenho em períodos de recessões em comparação com expansões.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

The main goal of this dissertation is to investigate the impact of socially responsible investing on the performance of fixed-income portfolios. According to the European Sustainable Investment Forum (EUROSIF, 2016, p. 9), "Sustainable and Responsible Investment is a long-term oriented investment approach, which integrates ESG (Environmental, Social and Governance) factors in the research, analysis and selection process of securities within an investment portfolio. It combines fundamental analysis and engagement with an evaluation of ESG factors in order to better capture long-term returns for investors, and to benefit society by influencing the behaviour of companies."

There is clearly a growing number of investors applying extra-financial criteria in their bond portfolio selection process. For instance, regarding socially responsible investment (SRI) asset allocation in Europe, there has been a significant increase in bond investing, which represented by December 2015 around 64% of the SRI market compared to 40% registered in December 2013. By 2015, the equity segment of the SRI market accounted for around 30% of SRI assets under management, indicating a significant decrease from the previous year's 50% weight (EUROSIF, 2016).

The growth of SRI in financial markets has attracted the interest of academics, who discuss the financial impact of considering social criteria in the portfolio selection process. Does considering social criteria benefit or penalize portfolio performance? Most studies in the financial literature address this issue by comparing the financial performance of SRI and conventional mutual funds, as Bauer *et al.* (2005) and Renneboog *et al.* (2008) in relation to equity investments, and Derwall and Koedijk (2009) and Leite and Cortez (2018) in relation to fixed-income investments. However, analyzing the effects of socially responsible investments by evaluating SRI mutual fund performance has some limitations. First, it is difficult to disentangle the performance that is due to the skills of the fund manager with the performance that is associated to the socially responsible characteristics of the companies held by the funds (Kempf and Osthoff, 2007). Second, fund returns reflect management fees and not only the returns of the funds' holdings (Auer, 2016). Third, there is evidence suggesting that the fact that a fund is classified as socially responsible does not ensure that it

is complying with ethical or social criteria (Auer, 2016). To avoid these problems, several studies propose evaluating the performance of socially responsible investments by analyzing the performance of synthetic portfolios formed on the basis of stocks' social characteristics. This dissertation follows this approach to evaluating socially responsible investments.

Despite the growth of asset allocation to socially responsible bonds, the majority of the empirical evidence on socially responsible investments is concentrated on the equity segment of the market and the performance of SRI fixed-income investments is far less explored. Furthermore, the few studies on the SRI fixed income area focus on the performance of actively managed mutual funds. In addition, most studies on the relationship between social ratings, credit costs, and default risk focus on the US market (Schröder, 2014). To the best of our knowledge, Hoepner and Nilsson (2017) is the only study exploring how the ESG ratings of companies can affect the performance of fixed-income portfolios, but their study is limited to the US market. In relation to the European market, to the best of our knowledge, there is no investigation on the performance of SRI bond portfolios formed on the basis of social criteria.

In this context, the main objective of this dissertation is to contribute to the SRI portfolio literature by investigating the effects of considering social criteria in European fixed income portfolios. In particular, we form mutually exclusive bond portfolios with high and low social ratings and investigate the impact of several social screens on the performance of such bond portfolios. Besides an aggregate measure of corporate social responsibility (CSR), we also consider individual dimensions of CSR, because there might be important social characteristics of companies that might end up hidden when using a broad indicator of CSR, as argued by Hoepner *et al.* (2016). The sample consists of Eurozone companies covered by ASSET4 ESG between 2003 and 2016. This source of social data has been used in several recent studies on SRI portfolio performance. Some of the advantages of ASSET4 are the consistency in the reporting and a lower likelihood of occuring matching errors, since Thomson Reuters also provides financial information on the companies (Gonenc and Scholtens, 2017). We employ the positive/ESG integration and best-in-class strategies in the bond portfolio selection process, as in Kempf and Ostoff (2007). However, since socially responsible investments following the ESG integration approach represent a higher amount of assets under management (AuM) in 2015 than those using the best-in-class approach, this

study focuses mainly on portfolios formed on the basis of the ESG integration strategies.[1] Portfolio performance is evaluated using unconditional and conditional multi-factor models with both time-varying alphas and betas. In addition, we also present empirical evidence on time-related issues, since several studies seem to suggest time-dependency of socially responsible portfolio performance. In terms of bond investing, to the best of our knowledge this is the first study providing evidence that supports the errors-in-expectations hypothesis and the shunned-stock hypothesis.

The results of this study are of major interest for academics and investors. The investment decision can be viewed as a top-down process: (i) Capital allocation between the risky portfolio and risk-free assets, (ii) asset allocation in the risky portfolio, and (iii) security selection (Bodie *et al.*, 2014, p. 205). The investor's degree of risk aversion is used to calculate the optimal proportion of the capital allocation between the risky portfolio and the risk-free asset. In addition, since risky portfolios of less than perfectly correlated asset classes offer diversification benefits, the incorporation of equity securities and bonds in a portfolio increases efficiency gains. As mentioned by Leite and Cortez (2018), fixed-income studies lead to better asset allocation decisions because they improve the understanding of socially responsible investments for a different asset class besides equity. Therefore, it is useful to investigate whether it is possible to do "well while doing good" with fixed-income portfolios. However, according to modern portfolio theory, the imposition of any social screens will restrict the investors' investment universe, leading to lower portfolio performance. Therefore, the empirical results on the effects of social screening in portfolio performance has implications to controversial issues in Finance, such as market efficiency and asset pricing. Moreover, although several studies claim that bonds constitute a homogenous asset class and bond returns are a function of systematic risk factors (Derwall and Koedijk, 2009), other studies (e.g., Dynkin *et al.*, 1999; Hottinga *et al.*, 2001; and Dynkin *et al.*, 2002), show that the firm-specific risk is also a share of the risk of corporate bonds. This suggests that SRI might also have an impact on bond investment performance.

The remainder of this paper is organized as follows. Chapter 2 reviews the literature on the performance of SRI funds and portfolios. Chapter 3 presents the methodology used to

---

[1] According to EUROSIF (2016), all responsible investment strategies have grown in the last years in Europe. By 2015, investments following the ESG Integration represented €1 900 billion AuM, whereas the best-in-class approach represented €493 billion AuM.

assess portfolio performance and chapter 4 describes the data. Chapter 5 presents and analyzes the empirical results. Finally, chapter 6 presents the main conclusions, the limitations of this study, as well as suggestions for future research.

2. LITERATURE REVIEW

This chapter reviews the literature on the impact of non-economic indicators on portfolio performance. It starts with an overview of socially responsible investing and the performance of equity and fixed-income SRI portfolios. Although this dissertation follows the approach that evaluates the performance of socially responsible investments by forming synthetic SRI portfolios, the discussion will also cover studies on indices and actively managed mutual funds, considering the attention the literature has given to this area. It is also important to keep in mind that studies that evaluate the performance of synthetic SRI portfolios may form portfolios according to an aggregate indicator of CSR or, alternatively, using individual dimensions of CSR (e.g., environment, governance, labour relations, etc.). Several studies (e.g., Hoepner *et al.*, 2016) find that the relation between financial performance and social performance is better assessed when using indicators that represent individual dimensions of CSR, such as environment, labour relations, and governance. Hence, this chapter distinguishes portfolio studies that use aggregate measures of CSR from those that use such individual dimensions of CSR as indicators of the companies' social standards. Finally, this chapter discusses empirical evidence on time-related issues, since several studies seem to suggest time-dependency patterns of socially responsible portfolio performance.

2.1. THE PERFORMANCE OF SOCIALLY RESPONSIBLE INVESTMENTS: AN OVERVIEW

There are several hypotheses on the performance of socially screened portfolios. The hypothesis of underperformance of SRI portfolios derives directly from finance theory. According to Markowitz (1952), the portfolio mean-variance optimization framework implies diversification across companies with different economic and financial characteristics. Accordingly, the imposition of any social screens will restrict the investment opportunity set, leading to lower portfolio performance. The additional information costs implied by the screening process are another argument in favour of a lower performance of SRI portfolios (Cortez *et al.*, 2009). The underperformance hypothesis of SRI portfolios is further supported by several studies (e.g., Hong and Kacperzyck, 2009; Statman and Glushkov, 2009; Derwall *et*

al., 2011) that find that stocks in controversial activities such as tobacco, alcohol and weapons, which are typically avoided by values-driven investors, yield abnormal returns. By excluding these stocks, socially responsible investors will not be able to benefit from these abnormal returns, whereas other investors might do so.

In contrast, consistent with stakeholder theory (Freeman, 1984), supporters of the outperformance of socially responsible investing argue that companies that consider the all stakeholders' interests might benefit from long-term higher performance. The underlying argument is that companies that are more socially responsible are better managed - CSR can be viewed as a source of competitive advantage that could be reflected in higher financial performance and higher shareholder value (Nollet *et al.*, 2016). This perspective is opposed to the traditional Milton Friedman (1962) viewpoint that the social responsibility of a business is solely to increase profits and any use of resources with a different purpose (such as CSR activities) will generate inefficiencies. A seminal paper on the benefits of socially responsible investing is Moskowitz (1972), who states that individuals and institutions may benefit from considering social screens in making investment decisions, and corporate actions on this matter have a positive influence on publicly traded shares.

2.1.1 SOCIALLY RESPONSIBLE INVESTING IN EQUITIES

More than four decades after Moskowitz's (1972) seminal paper, many empirical studies have addressed the performance of socially responsible investing. Some studies investigate this topic by evaluating the performance of socially responsible indices. Most studies in this line of research, such as Statman (2006), Schröder (2007) and Collison *et al.* (2008) document that the performance of social indices is comparable to the performance of conventional market indices.

Another line of research focuses on the performance of SRI mutual funds. Most empirical evidence on equity mutual funds finds no evidence of significant differences between the performance of ethical and conventional funds. Hamilton *et al.* (1993), using a US sample of funds, show that the performance of SRI funds is similar to the performance of

conventional funds. However, it is important to mention that the authors only use the traditional Capital Asset Pricing Model (CAPM) - based single-index model to evaluate fund performance. Reyes and Grieb (1998) and Bello (2005) document similar results for the same market. Studies on other markets include Scholtens (2005) on Dutch funds, Gregory and Whittaker (2007) on UK funds, Bauer *et al.* (2007) and Ayadi *et al.* (2016) on the Canadian market and Leite *et al.* (2018) on the Swedish market. The results of these studies are consensual in the sense that they find that the performance of SRI funds is not statistically different from that of conventional funs.

Papers that evaluate the performance of SRI funds in multiple markets also show that the performance of SRI funds is comparable to the performance of conventional funds (e.g., Schröder, 2004; Kreander *et al.*, 2005). Bauer *et al.* (2005), using a sample of German, UK and US SRI mutual funds for the period between 1990 and 2001, find no evidence of significant differences in performance between SRI and conventional funds, even after considering management fees. The authors use the Carhart (1997) 4-factor model and show that SRI funds are less exposed to the market portfolio and are either more growth-oriented or less value-oriented in comparison to conventional funds. The authors also find that conventional indices are more useful than ethical indices in explaining SRI fund returns. Cortez *et al.* (2009) investigate SRI equity and balanced mutual funds from seven European countries investing globally or in the European market and find that their performance is neutral relative to the market. As Bauer *et al.* (2006, 2007), the authors use the conditional approach to performance evaluation and show evidence of time-varying betas in SRI funds, suggesting that the conditional models are more reliable to evaluate performance than unconditional ones.

In contrast to the previous literature, a few studies find statistically significant differences between socially screened and unscreened mutual funds. Renneboog *et al.* (2008) analyze SRI funds from 17 countries between 1991 and 2003 also using conditional models. While the SRI funds' alphas are in almost all countries lower than those of conventional funds, their risk-adjusted returns are only statistically different in France, Ireland, Sweden, and Japan. On the contrary, Gil-Bazo *et al.* (2010) find that US SRI funds perform better than conventional funds.

As mentioned previously, studies on SRI mutual funds suffer from several limitations. First, SRI mutual fund performance can be a results of both the skills of the fund manager and the socially responsible characteristics of the companies held by funds (Kempf and Osthoff, 2007). Second, fund returns can reflect management fees, as also pointed out by Kempf and Ostoff (2007). Third, there is evidence suggesting that the fact that a fund is classified as socially responsible does not ensure that it is following the ethical or social criteria disclosed in their prospectus (Auer and Schuhmacher, 2016). For instance, Wimmer (2013) finds that funds classified as socially responsible considerably change their social standards over time (they tend to decrease after two years). Also, Utz and Wimmer (2014) document that, on average, SRI funds do not hold more socially responsible companies than conventional funds do. To avoid these problems, several studies construct synthetic portfolios on the basis of an aggregate or a specific dimension of CSR and evaluate the financial performance of these portfolios. In this section, we will review studies that consider an aggregate indicator of CSR in forming SRI portfolios. Aggregate indicators of CSR reflect the multidimensional nature of this concept, and as Waddock and Graves (1997, p.304) mention, encompass a wide range of aspects: inputs (e.g., investment in environmental equipment), internal processes (e.g., treatment of women and minorities, nature of products, relationship with customers) and outputs (e.g. community relations). Waddock *et al.* (2000) further claim that aggregate measures of CSR represent a significant advance over prior unidimensional measures.

According to Kempf and Osthoff (2007), one of the most cited studies in the field, investors can earn high abnormal returns from a strategy involving a long position in the high-rated social portfolio and a short position in the low-rated social portfolio. This study considers a sample of US companies over the period from 1991 to 2004 using ESG ratings provided by KLD. This conclusion is even more clear considering industry-balanced investment portfolios, i.e., using the best-in-class screening strategy, since this approach typically leads to the highest alphas (up to about 8.7% per year), specially with extreme social ratings. The results show that the alphas remain significant even after taking into account reasonable transaction costs. Furthermore, the results obtained from value- and equally-weighted portfolios are similar.

Filbeck *et al.* (2009) also find a positive effect of considering SRI screens on portfolio performance. The authors evaluate the performance of a portfolio consisting of the *100 Best*

Corporate Citizens published by *Business Ethics* compared to the performance of the S&P 500 and that of a matched sample of companies (matched by market capitalization and BE/ME ratio). The authors find that an investor can form a portfolio of the new companies on the list and outperform both benchmarks in the long term. For this purpose, investors can rebalance their portfolio holdings every year, by adding the newly listed companies and dropping the consecutive winners. The authors use several approaches to evaluate portfolio performance, such as the Fama and French (1993) three-factor and the Carhart (1997) four-factor models.

Borgers *et al.* (2013) develop an aggregate stakeholder-relations index on a yearly basis based on social indicators provided by KLD. The authors state that a portfolio composed of the top-ranked social stocks earn a higher average annualized risk-adjusted return than its bottom-ranked counterpart, which is consistent with the two previous studies. In turn, Mollet *et al.* (2013) analyze a sample of small European growth SRI firms selected according to the social scores provided by the sustainability research of the Zurich Cantonal Bank and find that the portfolio yields statistically significant positive abnormal returns.

Auer (2016) argues that the type of ethical screening strategy has implications on portfolio performance. Negative screens with low cut-off rates that exclude unrated stocks provide a significantly higher performance than a passive benchmark strategy. In contrast, positive screens can cause portfolios to underperform the benchmark because of a loss of diversification. Therefore, investors should eliminate the worst socially rated firms.

Several studies find a zero effect of social screens in portfolio performance. Statman and Glushkov (2009), using a US sample over the period from 1992 to 2007 from KLD database, find evidence that stocks of socially responsible companies yield higher returns than those of conventional companies. However, this advantage is offset from the exclusion of shunned companies. Hence, they support the "no effect" hypothesis. Similar to Kempf and Osthoff (2007), the authors also state that investors should consider best-in-class portfolios since this strategy leads to the best performance. Unlike Kempf and Osthoff (2007), the authors find differences in the statistical significance of the excess returns between value- and equally-weighted portfolios. They attribute this variation to the higher standard deviation of the returns of value-weighted portfolios. In consequence, these portfolios are less diversified.

According to Halbritter and Dorfleitner (2015), there are no significant return differences between portfolios of companies with high and low ESG rating levels. This finding is robust for a variation of portfolio cut-offs, weighting schemes and also when considering industry-balanced investment portfolios. The authors state that investors should no longer expect abnormal returns on this matter. The sample includes ESG data of ASSET4, Bloomberg and KLD for the US market from 1991 to 2012.

Outside the US, Van de Velde *et al.* (2005) show that portfolios of high-rated European companies perform better than low-rated ones but not to a significant extent. Brammer *et al.* (2006) suggest a negative effect of SRI screens in the UK since they find that companies with low social ratings present higher abnormal returns than companies with high social ratings. However, the differences between the two portfolios are not statistically significant. Brzeszczyñski and McIntosh (2014) suggest a positive effect of SRI screens in the UK market, but the results are not statistically significant.

Mollet and Ziegler (2014) address the US and European market and document that socially responsible stocks are correctly priced by market participants since there are no statistically significant abnormal returns resulting from SRI portfolios. The authors suggest that the learning process of market participants in the years before 1998 eliminated possible errors-in-expectations of investors and, therefore, the market has recognized the positive impact of CSR.

Auer and Schuhmacher (2016) document that investors in the US and Asia-Pacific market tend to have no significant financial advantages or disadvantages from investing in an (un)ethical portfolio choice with high- or low-rated stocks. However, in certain industries and depending on the ESG criterion, investors in Europe pay a price for being socially responsible in their stock selection process.

Galema *et al.* (2008) explain the discrepancy between the theoretical literature that suggests a positive relationship between SRI and stock returns and most of the empirical literature that finds a zero effect of SRI on performance. On the one hand, the impact of aggregated ESG scores can generate misleading inferences (Callan and Thomas, 2009) and eliminate any relationship if individual dimensions of social responsibility have opposite

effects on performance. However, the authors find little evidence in support of this hypothesis. Derwall *et al.* (2011) also put forward an explanation for the neutral performance of SRI portfolios. Derwall *et al.* (2011) develop two hypotheses with contrasting effects on performance, which ultimately will cancel out each other and result in neutral portfolio performance. The shunned-stock hypothesis claims that values-driven investors shun socially controversial stocks and, hence, these stocks present higher returns. Alternatively, according to the errors-in-expectations hypothesis, socially responsible stocks might have abnormal returns if the market does not immediately incorporate CSR practices on the companies' intrinsic value. However, the authors expect that any abnormal returns associated to errors-in-expectations to be temporary, since in the long run it is expected that investors recognize the importance of CSR information.

According to Galema *et al.* (2008), it is also relevant to mention that most researchers control for risk using the high minus low (HML) book-to-market ratio from the Fama and French (1992) asset pricing model. Nevertheless, for firms with equal risk levels, socially responsible stocks present lower book-to-market ratios because of their excess demand. As a result, the alphas do not capture CSR effects. In addition, Schröder (2014) documents that funds and indices do not generate outperformance in most cases because they include the few companies with very good social ratings and companies with not so good ratings. Indeed, Kempf and Osthoff (2007) only find a significant outperformance for portfolios formed with extreme social ratings (the 10% companies with the best rating). Using the top 50% of social ratings, the results cease to be statistically significant (Kempf and Osthoff, 2007).

A set of studies provide evidence of the effects of SRI on performance by analyzing how sin stocks perform. Sin stocks are appealing to analyze due to their addiction and undesirable consequences (Hong and Kacperczyk, 2009). Hong and Kacperczyk (2009) define sin stocks as publicly traded companies involved in the production of alcohol, tobacco, and gaming. Considering the market segmentation that stems from Merton's (1987) incomplete information model, the authors present two reasons for sin stocks being cheaper than conventional stocks. The model posits that neglected stocks tend to trade at a discount because they have a smaller investor base. Under this reasoning, Hong and Kacperczyk (2009) argue that the avoidance of sin stocks by investors leads to lower prices relative to their fundamental values because of limited risk sharing. Also, Merton (1987) shows that the CAPM

does not holds and unsystematic risk matters under limited risk sharing. Hence, the increased litigation risk should increase the expected returns of sin stocks. Hong and Kacperczyk (2009) focus on sample of US companies from 1926 to 2004 and show that sin stocks outperform comparable stocks even after considering robustness tests and the extension of the analysis to international markets as an out-of-sample test.

The outperformance of sin stock portfolios is also documented in other studies. Liston and Soydemir (2010) find statistically significant positive and negative abnormal returns for a sin portfolio and for a faith-based portfolio (using religious screens), respectively. The authors point out the norm-neglect hypothesis of Hong and Kacperczyk (2009) as a possible explanation for the sin stock portfolio outperformance. Trinks and Scholtens (2017) also find that investing in controversial stocks can result in additional risk-adjusted returns. The authors investigate the performance of an international sample of companies on fourteen potentially controversial issues over the period between 1991 and 2012.

Kempf and Osthoff (2007) and Statman and Glushkov (2009) also present somewhat similar conclusions, but the differences between sin and accepted stocks are not statistically significant.[2] This can be due to the wider definition of controversial business areas from the KLD database (Derwall *et al.,* 2011) since it also includes firms with firearms, military, and nuclear operations. Fabozzi *et al.* (2008), using a sample of 21 markets between 1970 and 2007, find a superior performance of the sin portfolio, which is robust across different time periods, industries, and national markets. The authors consider a wider range of controversial industries, which also includes biotech and adult services.

Lobe and Walkshäusl (2016) define sin stocks as companies involved in the adult entertainment, alcohol, gambling, nuclear power, tobacco, and weapons industries, according to the Industry Classification Benchmark (ICB). The authors do not find evidence that sin stocks or socially responsible stocks outperform the market in different regions over the period 1995 to 2007. Considering that Hong and Kacperczyk (2009) show evidence of positive abnormal returns, Lobe and Walkshäusl (2016) also use an approach that is similar to the one used by Hong and Kacperczyk (2009) for the US sample. However, their first

---

[2] Statman and Glushkov (2009) find a positive alpha when using the CAPM and the three-factor model but this result is not observed when using the four-factor model.

conclusions still hold. An as explanation, the authors state that the long sample period of Hong and Kacperczyk (2009) can drive their results as it appears that the 1960s and 1970s mainly drive the US financial outperformance of sin stocks.

The studies reviewed in this chapter present different conclusions regarding the relationship between social and financial performance. In order to shed light on the effects of including social criteria in the portfolio selection process, Revelli and Viviani (2015) perform a meta-analysis of extant studies on SRI performance. Focusing on empirical studies using data on mutual funds, portfolios and indices between 1972 and 2012, the authors conclude that the incorporation of CSR in the stock market does not add financial costs nor benefits compared with conventional investments. However, the review study of Schröder (2014), calls attention to the possibility that long-short portfolios outperform only when using extreme social ratings, as mentioned previously. In addition, the author points out the outperformance portfolios of stocks in sin sectors.

2.1.2   SOCIALLY RESPONSIBLE INVESTING IN BONDS

As previously mentioned, there are fewer studies on socially responsible investing in bonds compared to equities. Regarding the empirical evidence on fixed income mutual funds, Goldreyer and Diltz (1999) find negative and statistically significant alphas for US SRI bond funds. Derwall and Koedijk (2009) evaluate the performance of SRI fixed-income mutual funds in the US from 1987 to 2003. Using multi-index performance evaluation models, the authors show that SRI bond funds perform similar to conventional funds, while SRI balanced funds outperform conventional funds.

More recently, Henke (2016) shows that over the period 2001 to 2014 SRI bond funds in the US and the Eurozone outperform not only the market (however only at the 10% level in the Eurozone market) but also conventional funds. Furthermore, Henke (2016) shows that the outperformance of SRI bond funds is more evident during crisis periods. By analyzing the holdings of the funds, the author concludes that fund managers appear to implement a worst-

in-class exclusion rather than a best-in-class inclusion screening typically more used by SRI equity funds.

Leite and Cortez (2018) investigate the performance of SRI bond and balanced funds domiciled in the leading Euro-area markets (France and Germany) from 2002 to 2014. The authors find no differences in the performance between SRI balanced and conventional funds. However, SRI bond funds significantly outperform their conventional peers, in line with Henke (2016). The authors claim that this outperformance seems more related to the government and not the corporate bonds they hold. Also, both SRI and conventional balanced funds as well as conventional bond funds underperform the market. These results are in constrast with those of Derwall and Koedijk (2009), who find that the performance of SRI balanced funds is neutral. Leite and Cortez (2018) also find that SRI bond funds exhibit neutral performance, while Henke (2016) observes outperformance of their US peers.

Regarding the direct construction of synthetic fixed-income portfolios, Hoepner and Nilsson (2017), using a US sample over the period from 2001 to 2014, explore the effect of responsible investing on bond performance. To the best of our knowledge, Hoepner and Nilsson (2017) is the only study that explores if a trading strategy in bonds, based on their ESG ratings, leads to abnormal returns. The results of Hoepner and Nilsson (2017) do not show evidence that bonds of companies with high ESG ratings outperform those of low ESG ratings companies, in contrast with Kempf and Osthoff (2007) in the context of ESG equity portfolios. In fact, the authors suggest "no news is good news" in ESG bonds since bonds issued by companies with no strengths, concerns or controversies outperform by 1.12% per year in comparison to bonds issued by companies with strengths and concerns. This result is robust when controlling for industry differences, differences in remaining maturity, and when analyzing value- and equally-weighted portfolios.

There is also empirical evidence on the effect of CSR on credit spreads, cost of debt, and credit risk. Oikonomou *et al.* (2014), using a US sample between 1991 and 2008 and social rating provided by KLD find that engaging in CSR and avoidance of controversies reduces the corporate bond yield spread and thus decreases the cost of corporate debt. The authors also find that higher levels of CSR lead to better credit quality and lower perceived credit risk. Ge and Liu (2015) and Bauer and Hann (2010) find similar results, with the latter concentrating

on environmental information. Additionally, according to Oikonomou *et al.* (2014), CSR seems to be important in the cases of highly rated bonds (A+ to A−) or very low-rated bonds (CCC+ or lower). Since low credit rated bonds present high yields, issuers of speculative grade bonds can benefit the most in absolute terms from the reductions in the cost of debt from CSR practices (Oikonomou *et al.,* 2014). However, Menz (2010) shows, for a sample of European companies, that socially responsible firms do not have lower cost of public debt. In addition, the authors argue that a possible explanation might be that CSR information might not yet been incorporated into bond valuations.

Drut (2010) investigates socially responsible performance by analyzing sovereign bonds of 20 developed countries. The author finds that investors can create sovereign bond portfolios with high social ratings without significantly losing diversification and, hence, without a significant loss of mean–variance efficiency.

Schröder (2014) reviews several empirical studies on the performance of socially responsible investments and concludes that engaging in CSR practices reduces the default risk and the cost of debt. However, the author points out that most of the empirical evidence is limited to US companies.

2.2. THE EFFECT OF SPECIFIC CSR DIMENSIONS ON PORTFOLIO PERFORMANCE

Besides studies that construct portfolios based on an aggregate dimension of CSR, there are others that investigate the relationship between specific dimensions of CSR and financial performance. The argument to use indicators that represent individual dimensions of CSR is that there might be important social characteristics of companies (e.g., environment, labour relations, and governance) that might end up hidden when using a broad indicator of CSR. The underlying idea is that firms that target the concerns of specific stakeholders might benefit from more sizeable effects of this behavior (Hoepner *et al.,* 2016).

This section discusses the literature that evaluates SRI performance by constructing synthetic portfolios on the basis of a specific dimension of CSR.[3]

Regarding the environmental criterion, Derwall *et al.* (2005) define eco-efficiency as an economic value that corresponds to the difference between what a company creates and the waste it makes (Derwall *et al.*, 2005, p. 52). Using a US sample of companies over the 1995-2003 period and based on Innovest Strategic Value Advisors' corporate eco-efficiency scores, the authors conclude that the high-rated eco-efficient portfolio not only shows statistically significant abnormal returns but also outperforms the low-rated eco-efficient portfolio. The results are also statistically significant after taking into account transaction costs, when considering industry-adjusted portfolio returns using sensitive sectors only, and when using a portfolio based on the best-in-class approach. However, the conclusions do not hold when using equally-weighted portfolios. Between the latent risk factors and the mispriced hypothesis, the authors point out the market's inability to price eco-efficiency in the short run as an explanation for the results. In fact, this conclusion might also explain the reduction of the eco-efficiency premium in environmentally sensitive sectors where the benefits of eco-efficiency are more obvious.

Kempf and Ostoff (2007), using environmental scores from KLD, report that the high-rated portfolio yields a significant positive alpha, but this result is not observed neither for the long-short strategy nor the best-in-class approach. Galema *et al.* (2008), including all stocks covered by KLD database, document that constructing portfolios on the basis of an environmental screening does not lead to a significant outperformance. Brammer *et al.* (2006) find that the difference between high- and low-rated portfolios is not statistically significant in the UK. In turn, Van de Veld *et al.* (2005), Statman and Glushkov (2009) and Auer (2016) find no statistically significant results for the environmental screen. Auer and Schumacher (2016) find that high-rated environmental portfolios significantly underperform their benchmarks in the consumption and miscellaneous sectors in Europe. However, the results are not statistically significant in the United States and in the Asia-Pacific region.

---

[3] For the sake of brevity, this section does not include studies that evaluate the performance of SRI funds that focus on a specific dimensions of CSR. Examples of such studies include Climent and Soriano (2011), Muñoz, Vargas and Marco (2014) and Silva and Cortez (2016) on funds that screen on environmental criteria (so-called green funds).

Considering fixed-income securities, Hoepner and Nilsson (2017) show that any environmental news of a company has a negative effect on performance since both the high- and low-rated portfolios underperform the market benchmark between 2001 and 2014.

Regarding the social criterion (strictly speaking)[4], Edmans (2011), using the list of "100 Best Companies to Work for in America", finds that over the period 1984 to 2009 firms with high levels of employee satisfaction generate abnormal returns, which is consistent with human relations theories. A value-weighted portfolio of these companies presents an alpha of 3.50% and the results do not change after considering firm characteristics, different weighting methodologies, industry-adjusted portfolios and the removal of outliers. The authors explain the existence of abnormal returns with the possibility that the market fails to immediately incorporate the value of CSR practices.

Carvalho and Areal (2016), using the same list of companies as Edmans (2011), find that a portfolio of companies listed as the best places to work has a monthly abnormal return of 0.44% over the period 1998 to 2010. However, when risk is allowed to vary over time, by applying conditional models with a dummy to distinguish different market states (bull and bear markets), the alpha is no longer statistically significant for the aggregated portfolio. The authors also show that abnormal returns are concentrated in the top scoring companies.

Kempf and Ostoff (2007) find that the community screen yields a significant positive alpha when using the long-short strategy and the best-in-class approach for equally- and valued-weighted portfolios. The employee relations screen yields a significant positive alpha when using the long-short strategy and the best-in-class approach for valued-weighted portfolios. However, when considering the equally-weighted portfolios, only the long-short strategy presents significant results. The authors also suggest a positive abnormal return on the diversity and human rights screens, but the results are not statistically significant.

Statman and Glushkov (2009), using the same database as Kempf and Ostoff (2007) and simulating a long-short strategy based on the best-in-class approach, find positive and statistically significant results for portfolios formed on the basis of the employee relations screen. However, the community screen yields a significant result only when considering

---

[4] The expression "social" is used in refer to the second criteria in the ESG expression ("(Social") and not to social criteria in a broad sense.

equally-weighted portfolios. Galema *et al.* (2008) do not find statistically significant alphas for the employee and community screens using the Carhart (1997) 4-factor model. None of the three previous studies present a significant result for the human rights and products screen. Van de Veld *et al.* (2005) suggest that companies with the best customer and supplier ratings outperforms. However, once again, the results are not statistically significant. Brammer *et al.* (2006) suggest that high scores on the community and employment indicators lead to lower and higher returns, respectively. However, the differences between high- and low-scoring portfolios in both criteria are not statistically significant. According to Auer (2016), the social criterion does not generate any significant additional value. Considering this criterion, Auer and Schumacher (2016) find that high-rated portfolios significantly underperform their benchmarks in the consumption and financial sectors in Europe. However, the results are not statistically significant in the United States and in the Asia-Pacific region.[5]

In the fixed-income area, Hoepner and Nilsson (2017) suggest that bonds issued by companies with no ratings or neutral ESG scores outperform for community, diversity, and product safety screens. After controlling for industry effects, the portfolios classified as "no ratings" for employee relations, human rights and product safety screens present significant positive alphas.

Regarding the governance criterion, Gompers *et al.* (2003) construct a Governance Index "as a proxy for the balance of power between managers and shareholders" (Gompers *et al.*, 2003, p. 109). The authors show that during the 1990s, an investment strategy going long in firms with strong shareholder rights and short in firms with weak shareholder rights earns abnormal returns of 8.5 percent per year. Focusing on the governance dimension, Auer (2016) states that a strategy of using negative screens with small cut-off rates provides a significantly higher performance than a passive benchmark strategy. However, Van de Veld *et al.* (2005), Galema *et. al* (2008) and Statman and Glushkov (2009) find no statistically significant results when using this criterion. Core *et al.* (2006) consider that the period studied in Gompers *et al.* (2003) may drive their results (Statman and Glushkov, 2009).

---

[5] Although this section is focused on synthetic portfolio studies, it is worth to mention that Renneboog *et al.* (2008) find that actively managed funds adopting the community screen provide better returns.

2.3. TIME EFFECTS ON THE PERFORMANCE OF SOCIALLY RESPONSIBLE INVESTMENTS

2.3.1 LONGEVITY OF SRI OUTPERFORMANCE

According to Derwall *et al.* (2011), the SRI movement can be divided into a values-driven approach, in which social and personal values are the sources to integrate CSR criteria, and into a profit-seeking approach which seek traditional financial performancee. The authors distinguish two hypotheses. The shunned-stock hypothesis, following Merton's (1987) incomplete information model, claims that values-driven investors shun socially controversial stocks and hence these stocks will generate higher returns. Alternatively, according to the errors-in-expectations hypothesis, socially responsible stocks might have abnormal returns if the market does not immediately incorporate the value of CSR on expected cash flows. The authors expect enduring positive abnormal returns of shunned-stock over time due to the universal nature of social values. In contrast, the authors expect any abnormal returns associated to errors-in-expectations to be temporary, since in the long run investors should incorporate CSR information into security valuations. By constructing two portfolios and evaluating its performance over the period from 1992 to 2008, the authors confirm these claims. The annualized abnormal return on the shunned-stock portfolio ranges from 2.58% to 2.86% and is always statistically significant, while the annualized abnormal return of a high-rated portfolio formed on the basis of the employee relations criterion decreases from 5.62% to 2.81% over time.

Borgers *et al.* (2013) further provide evidence that stakeholder information is mispriced due to the market's inability to recognize CSR practices in the short term. Consistent with the errors-in-expectations hypothesis, the risk-adjusted returns decline as investors' attention for stakeholder information increases over time. For a range of US firms over the period 1992 to 2009, the authors split the sample in two subperiods. During the first subperiod the authors find that average risk-adjusted return differences between high- and low-rated portfolios are positive and statistically significant. In contrast, during the second subperiod risk-adjusted returns are not significantly different from zero in most cases.

Edmans (2011) points out that the abnormal returns of companies with high levels of employee satisfaction decrease over time for two reasons associated to mispricing rather

than risk. On the one hand, employee satisfaction is not an enduring characteristic, since each year one-third of companies fall off the list of best companies to work for. Hence, these companies should generate smaller outperformance. On the other hand, even for others that remain on the list for several years, the mispricing may be corrected over time as the market learns about the value of CSR practices. Edmans (2011) confirms both hypotheses by calculating the cumulative abnormal returns (CAR) and buy-and-hold returns (BHAR) of the portfolios of the best companies to work for compared to other companies. The returns from BHAR decrease over time and become close to zero after five years. In order to test the second hypothesis, the author focuses on companies which continue on the list for at least the next five years. Since the returns decrease in the fifth year, the results are consistent with the mispricing narrative. In order to test the first hypothesis, the author considers portfolio I that is rebalanced and reweighted each year, portfolio II that is not rebalanced nor reweighted each year, and portfolio III that includes only companies that dropped from the list. Since the results show that portfolio II underperforms portfolio I, and portfolio III underperforms both portfolios, the author argues that returns decrease over time because satisfaction is not an enduring characteristic.

Mollet *et al.* (2013) present two hypotheses that can explain the positive abnormal returns of SRI portfolios. First, there are errors-in-expectations in the market, as hypothesized by Derwall *et al.* (2011). In fact, regarding the corporate governance dimension, Bebchuk *et al.* (2013) find evidence of abnormal returns from 1990 to 1999. However, the abnormal returns disappeared between 2000 and 2008 as a result of increased attention by the market to corporate governance (Mollet *et al.*, 2013). This situation is consistent with the learning hypothesis. Second, Mollet *et al.* (2013) state that an outperformance can stem from better than expected fundamentals of companies or from rising stock demand while the companies' fundamental value hold on. This is consistent with the growth in SRI assets under management having led to increased stock prices.

Halbritter and Dorfleitner (2015) suggest that investors should no longer expect abnormal returns by trading a SRI portfolio. Splitting the sample into three subperiods between 1991 and 2012, the authors find that the effect of CSR in performance declines over time. While in earlier years high-rated companies significantly outperformed their lower-rated counterparts, as of 2012 this outperformance disappears.

2.3.2 RESILIENCE IN SRI PORTFOLIO PERFORMANCE IN TIMES OF CRISIS

The issue of whether SRI portfolios perform better than conventional portfolios in periods of crisis is a relevant topic in the literature. There are theoretical arguments in favour a higher performance of SRI firms in times of turmoil. As Hoepner *et al.* (2016) argue, companies that are highly committed to CSR practices generate reputational wealth and relational capital that can prevent decline in value during market crises. This argument is supported by Lins *et al.* (2017), who find that during the 2008-2009 financial crisis, firms with high levels of CSR outperform firms with low levels of CSR. According to Lins *et al.* (2017), CSR can be viewed as an insurance policy that protects firms in periods where there is a general crisis of confidence in the economy.

The issue of whether SRI portfolios perform differently has been addressed at the mutual fund level. For instance, in relation to actively managed SRI equity funds, Nofsinger and Varma (2014), Becchetti *et al.* (2015) and Leite and Cortez (2015) find that SRI funds provide additional protection during periods of market crisis.

In terms of equity portfolio studies, Carvalho and Areal (2016) find that the performance of portfolios of the best companies to work for is not different in times of crisis compared to periods of non-crisis. In particular, the top companies of the list outperform the market in expansion periods and this performance does not change in times of crisis. The authors thus conclude that a portfolio of the best companies to work for is, thereby, resilient in times of crisis.

Turning to the fixed-income area, considering that companies with high levels of CSR will be less exposed to risk and that investors pay greater attention to risk in turbulent times, one would expect that any positive effect of social screening on bond portfolio performance would occur during crisis periods (Henke, 2016). Henke (2016) analyzes the performance of SRI bond funds considering recessions and expansion periods according to the National Bureau of Economic Research (NBER) business cycles. Alternatively, the author identifies bull and bear markets as proxies for periods of non-crisis and crisis, respectively. The results show that US and Eurozone SRI bond funds present a significant outperformance compared to conventional funds during recessions and bear markets. Therefore, Henke (2016) concludes

that the outperformance of SRI bond funds is especially driven by the performance in crisis periods. During non-recession periods, only the alphas of US SRI bond funds show a significant outperformance compared to their conventional counterparts. In bull market periods, there are no statistically significant results.

Leite and Cortez (2016) evaluate bond mutual fund performance over the 2002-2014 period. The model used includes two dummy variables in the 4-factor model, which allows to separate the recession and expansion coefficients with respect to alphas and the coefficients of the risk factors. The authors find that the performance of SRI and conventional bond funds from France and Germany is negative and statistically significant in expansion periods and neutral during recessions. UK bond funds underperform both in periods of expansions and recessions. During expansions both French and German SRI bond funds significantly outperform conventional funds, while UK bond funds perform similar to their conventional peers. During recessions, only German SRI bond funds continue to outperform their peers, thus suggesting they provide some additional protection during market turmoil.

With respect to portfolio studies, Hoepner and Nillson (2017) split the fixed-income sample into January 2001 to December 2007 and January 2008 to December 2014 to distinguish between pre- and post-crisis periods. Although the results are not reported in the paper, the authors mention that the general conclusion of "no news is good news" results stems entirely from the second half of the sample period, i.e., in times of market crisis.

3. METHODOLOGY

This chapter presents the methodology used in this investigation. It starts by explaining portfolios' construction using the calendar-time portfolio approach. It also presents the performance evaluation models that will be used, as well as the estimation process of the models.

3.1. PORTFOLIO FORMATION

This investigation uses the calendar-time portfolio method, as in Edmans (2011) and Carvalho and Areal (2016). This approach to evaluating long-term performance presents a solution to the cross-correlation problem which is associated to methods such as cumulative abnormal returns (CAR) or buy and hold abnormal returns (BHAR) (Fama, 1998). This issue is particularly important with events that occur simultaneously because there is correlation between the events (Carvalho and Areal, 2016). In this case the returns of bonds in each portfolio are measured at the same time.

Following Kempf and Ostoff (2007), we employ the positive/ESG integration and best-in-class approaches. Positive ESG Integration refers to asset managers considering ESG factors into traditional financial analysis (Eurosif, 2016). Following Hoepner and Nilsson (2017) and Kempf and Ostoff (2007), at the beginning of period t two value-weighted portfolios are formed based on the stocks' social scores at the end of period $t-1$. In order to timely reflect the changes of ESG scores that may occur throughout the year in Aseet4 ESG scores, we consider the score at the end of month $t-1$. Portfolios are formed with respect to individual dimensions of CSR (environment, social and governance dimensions) as well as to an aggregate (combined) measure of CSR. For each individual dimension as well as for the aggregate measure, the high- and low-rated portfolio consist of the top (above the median score) and bottom (below the median score) of all firms with bonds in each period, respectively. The long-short portfolio corresponds to the differences in returns between a high- and low-rated portfolio, thereby representing the performance of a strategy of going

long in a high-rated portfolio and short in a low-rated portfolio. As a robustness test, we intend to test different cut-off portfolios. These portfolios are rebalanced monthly to account for any new securities issues or redemptions. The best-in-class strategy is used to avoid industry biases (Kempf and Osthoff, 2007). By following this approach, investors choose companies based on the best ESG scores in each industry (Eurosif, 2016). We first divide the companies into ten industries based on the Industry Classification Benchmark (ICB).[6] The portfolios for each industry are formed as described above for the positive screening strategy. To combine the different industry portfolios into one portfolio, we weight them according to the bond market index industry weights.

In order to compute returns, we start by calculating the discrete rate of return of each bond using the following equation:

$$R_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1 \tag{1}$$

where $R_{i,t}$ is the discrete rate of return of bond i in month t, $P_{i,t}$ is the price including interest payments[7] of bond i in month t, and $P_{i,t-1}$ is the price including interest payments of bond i in month $t-1$. To construct value-weighted portfolios, the following equation is used:

$$R_{p,t}^{VW} = \sum_{i=1}^{N} \frac{MV_{i,t-1}}{\sum_{i=1}^{N} MV_{i,t-1}} R_{i,t} \tag{2}$$

---

[6] The best-in-class approach considering the ICB industries was also implemented by Hoepner and Nilsson (2017), for example.
[7] Total return data is preferred to using only bond prices because it takes into account capital gains and interest payments.

where $R_{p,t}^{VW}$ is the rate of return of portfolio p in month t, $MV_{i,t-1}$ is the market value of bond i at the beginning of month t, and $r_{i,t}$ is the discrete rate of return of bond i in month t. Fama and French (2008) state that some anomalies are not robust to different weighting schemes (Edmans, 2011), because small and big companies have a different influence in the value- and equal-weighted portfolio return. One of the advantages of capitalization weighting is that it considers the relative value of portfolio securities. In addition, a capitalization weighting method requires less rebalancing than other types and, therefore, investors take less transaction costs (Christopherson *et al.*, 2009). However, some managers prefer do not assume a significant position in one company (Christopherson *et al.*, 2009). Therefore, equal-weighted portfolios are also constructed as a robustness test:

$$R_{p,t}^{EW} = \frac{1}{N} \sum_{i=1}^{N} R_{i,t} \tag{3}$$

where $R_{p,t}^{EW}$ is the rate of return of an equally-weighted portfolio in month t, $P_{i,t}$ is the price including interest payments of bond i in month t, $P_{i,t-1}$ is the price including interest payments of bond i in month $t-1$, and N is the number of bonds in portfolio p.

3.2. UNCONDITIONAL MODELS OF PERFORMANCE EVALUATION

Jensen´s (1968) alpha measures the difference between the actual portfolio return and the expected risk-adjusted return. This measure is based on the CAPM and uses a market benchmark as the unique risk factor. Although the single factor performance was initially applied to stocks, it has also been applied to bond securities (Silva *et al.*, 2003). The single-factor alpha is computed as:

$$r_{p,t} = \alpha_p + \beta_{1p} r_{m,t} + \varepsilon_{p,t} \tag{4}$$

where $r_{p,t}$ is the excess return of portfolio p in month t, β_{1p} is the systematic risk of the portfolio, $r_{m,t}$ represents the market's excess return in the same month and $\varepsilon_{p,t}$ is the error term, which has an expected value of zero.

As a single-factor model, only one source of systematic risk is priced in the CAPM. Considering extensive evidence that there are other risk factors that are priced, investment strategies based on those factors can easily show significant abnormal return estimates without meaning that fund managers are skillful. Given this limitation, and motivated by the development of the Arbitrage Pricing Theory (APT) by Ross (1976), that proposes a linear relationship between expected return and multiple sources of systematic risk, the mutual fund literature turned to multi-factor models of performance evaluation. Several studies, such as Fama and French (1992) and Carhart (1997) in the case of equity portfolios and Elton *et al.* (1995) in the case of fixed-income portfolios, show that multi-factor models can provide better descriptions of security returns. In the spirit of Elton *et al.* (1995) and Derwall and Koedijk (2009), the following multi-factor model is used:

$$r_{p,t} = \alpha_p + \beta_{1p} Bond_t + \beta_{2p} Default_t + \beta_{3p} Option_t + \beta_{4p} Equity_t + \varepsilon_{p,t} \tag{5}$$

where $r_{p,t}$ represents the excess returns of portfolio p in month t; $Bond_t$ represents the excess returns of the bond market index, thus capturing the exposure to investment-grade bonds; $Default_t$ is the return spread between a high-yield bond index and a government bond index and it is included to capture default risk compensation; $Option_t$ is the return difference between a mortgage-backed securities index and a government bond index; $Equity_t$ represent the excess returns of the equity market index, and $\varepsilon_{p,t}$ is the error term.

3.3. CONDITIONAL MODELS OF PERFORMANCE EVALUATION

The unconditional performance measures presented before do not take into account information about variations in the state of the economy or financial markets. This implies these models do not measure expected excess returns correctly when portfolios' risk exposures and risk premiums change over time (Christopherson *et al.*, 2009). In response to this concern, Ferson and Schadt (1996) propose a conditional performance evaluation model that assumes a linear functional form for the changing conditional beta of a portfolio, given a set of public information variables that proxy for the state of the economy. In this model, the conditional beta is defined as:

$$\beta_p(Z_{t-1}) = \beta_{0p} + \beta_p' \, z_{t-1} \qquad (6)$$

where $z_{t-1} = Z_{t-1} - E(Z)$, corresponding to the deviations of Z_{t-1} from their average values, β_{0p} represents the unconditional average beta, and β_p' measures the sensitivity of the conditional betas to the deviations of Z_{t-1} from their means. The incorporation of the information variables into Jensen's (1968) measure leads to the following regression, which is also known as the partial conditional model:

$$r_{p,t} = \alpha_p + \beta_{0p} \, r_{m,t} + \beta_p' \left(z_{t-1} \, r_{m,t} \right) + \varepsilon_{p,t} \qquad (7)$$

where α_p represents the conditional performance measure, that will take the value of zero if portfolio managers uses only publicly available information contained in Z_{t-1}. According to Ferson and Qian (2004), the conditional alpha measures the fund's excess return over a strategy that mimics the fund's risk dynamics over time based on the predetermined information variables. In contrast, the unconditional performance measures only compare a fund's return with benchmarks that present unchanged average exposures to risk over time.

Consistent with the semi-strong market efficiency described by Fama (1970), portfolio managers that use only publicly available information do not add value and, hence, do not generate a positive conditional alpha (Ferson and Qian, 2004). According to Christopherson *et al.* (2009), there are two components of portfolio managers' skills: market-timing ability using available macroeconomic information and the ability to pick stocks. Bond fund managers are more market timers than security pickers so they mainly adjust portfolio's duration to expected interest rates. (Silva *et al.*, 2003; Leite and Cortez, 2018). If they expect an increase of interest rates, they will decrease duration (and vice versa). Hence, it is important to employ conditional models because duration is linked to beta (Silva *et al.*, 2003).

The public information variables are 1-period lagged because it is expected that the manager's ability to exploit available macroeconomic information be reflected on portfolio positions only in the next period.

The partial conditional model of Ferson and Schadt (1996) allows beta to be time-varying and provides more reliable estimates of alpha relative to unconditional models (Christopherson *et al.*, 2009). However, the alpha remains constant. Christopherson *et al.* (1998) develop the full conditional model to allow for time-varying conditional alphas. In this model alpha is also a linear function of Z_{t-1}:

$$\alpha_p(Z_{t-1}) = \alpha_{0p} + A'_p z_{t-1} \tag{8}$$

where α_{0p} is an average alpha and the vector A'_p measures the response of the conditional alpha to the information variables. Consequently, this conditional model can be expressed as:

$$r_{p,t} = \alpha_{0p} + A'_p z_{t-1} + \beta_{0p} r_{m,t} + \beta'_p \left(z_{t-1} r_{m,t} \right) + \varepsilon_{p,t} \tag{9}$$

The extension of this conditional single factor model to a multifactor context considering the risk factors included in equation (5) is straightforward, and results in the following equation:

$$
\begin{aligned}
r_{p,t} = \ & \alpha_{0p} + A'_p z_{t-1} + \beta_{1p} Bond_t + \beta'_{1p}(z_{t-1} Bond_t) + \beta_{2p} Default_t \\
& + \beta'_{2p}(z_{t-1} Default_t) + \beta_{3p} Option_t + \beta'_{3p}(z_{t-1} Option_t) \\
& + \beta_{4p} Equity_t + \beta'_{4p}(z_{t-1} Equity_t) + \varepsilon_{p,t}
\end{aligned}
\tag{10}
$$

For this model to be implemented in practice, it is necessary to choose the public information variables to be used. Silva *et al.* (2003) analyze the ability of a set of variables in predicting bond returns in the European market. They consider three of the variables used by (Ilmanen, 1995), namely the term spread, real bond yield and inverse relative wealth, plus a dummy variable for the month of January. However, the authors do not to include the real bond yield variable in the bond fund performance evaluation model because it is highly correlated with the term spread[8] and it does not contribute significantly to the explanatory power when testing predictability of European bond returns. The results of this study show evidence that the term spread, inverse relative wealth and a January dummy are useful predictors variables of European bond excess returns.

The term spread is the difference between the yields of a long-term and a short-term bond and is a proxy for the expected bond risk premium. Theoretically, one expects a positive coefficient of the term spread, since expected returns on corporate bonds have a positive relationship with the variation in maturity premiums (differences between the expected returns on long- and short-term bonds). Company defaults and credit rating downgrades are more frequent during economic recessions and, therefore, it is possible to observe an increase on expected returns. Fama and French (1989) claim that the variation of the term spread is related to business cycles since it tends to be low near business-cycle peaks and high

---

[8] Near multicollinearity is a regression assumption violation that occurs when two or more independent variables are highly but not perfectly correlated with each other. Although this violation does not affect the consistency of the OLS estimates of the regression coefficients, the consequences are reflected in inflated OLS standard errors. Consequently, t-tests on the coefficients have little power.

near troughs (as identified by the NBER). However, this variable is not a perfect proxy for the expected bond risk premium because it also contains information about expected nominal or real rates (Ilmanen, 1995; Silva *et al.,* 2003). Indeed, Silva *et al.* (2003) find a negative coefficient of the term spread, contrary to what is usually observed for the US market. The fact that expectations about inflation rate are positively correlated with business cycles and consequently negatively correlated with expected returns might be a possible explanation for this result, as mentioned by the authors.

The inverse relative wealth (IRW) is used as a proxy for time-varying risk aversion. Ilmanen (1995) states that investors are more risk averse when their wealth is lower than their past wealth. Consequently, investors demand a higher compensation to take risky assets during recession periods. Following Ilmanen (1995) and Silva *et al.* (2003), the IRW variable is defined as:

$$IRW = \frac{ewa\ W_{t-1}}{W_t} = \frac{(W_{t-1} + 0.9 * W_{t-2} + \ 0.9^2 * W_{t-3} + \cdots) * 0.1}{W_t} \qquad (11)$$

where W_t is the real level of stock market at time *t,* and $ewa\ W_{t-1}$ is the exponentially weighted average of real stock market levels up to $t - 1$. The proxy for aggregate wealth is the real stock market index. Stock markets are a small share of the world wealth but they are probably the most volatile segment and are positively correlated with other segments of wealth (Ilmanen, 1995; Silva *et al.,* 2003). The coefficient of this variable is expected to be positive since risk aversion measured by the IRW is positively correlated with expected bond returns.

Finally, the dummy variable takes a value of 1 if the next month is the month of January and 0 otherwise. The goal is to capture a January seasonality, since there is probably an increased risk at this time of the year (Silva *et al.,* 2003).

As mentioned before, public information variables are 1-period lagged. However, expected returns might be dependent over time and, consequently, information variables as

lagged variables might show a better prediction of returns than they really do (Ferson *et al.*, 2003). If the information variables have large autocorrelations, this might lead to spurious relations. To reduce this possibility, Ferson *et al.* (2003) propose a stochastic detrending procedure which consists on the transformation of the lagged variable by subtracting a 12-month moving average of its own monthly observations. In addition, mean-zero variables are used in order to minimize possible scale effects on the results.

An alternative approach to condition fund performance to the economy involves using dummy variables to distinguish different market states, as in Moskowitz (2000), Kosowski (2006) and Areal *et al.* (2013). Whereas the conditional model of Christopherson *et al.* (1998) conditions performance and risk to the state of the economy by means of public information variables, the dummy variable model conditions performance and risk to different market states, such as bull and bear markets or expansion and recession periods. Hence, the latter model allows to verify if there are statistically differences on the performance and risk of SRI portfolios during different market regimes, i. e., during "good times" and "bad times".

The use of dummies to capture different market states avoids the assumption that betas and alphas are a linear function of the information variables (Areal *et al.,* 2013; Ferson and Qian, 2004; Silva and Cortez, 2016). The models of Ferson and Schadt (1996) and Christopherson *et al.* (1998) assume that a fund manager responds linearly to public information. However, the relationship between conditional risk and the public information variables is likely to be nonlinear (Ferson and Qian, 2004). Consequently, the regressions are likely to be biased and inefficient. Rather than dummy variables associated to different market states (e.g., crisis versus non-crisis), Ferson and Qian (2004) use dummy variables associated to the public information variables. In particular, they define a higher-than-normal state variable and a lower-than-normal state variable for each public information variable. The use of these dummy variables can provide more reliable estimates of alpha in cases where nonlinearity is important (Ferson and Qian, 2004).[9]

---

[9] Nevertheless, Areal *et al.* (2013) argue that the information variables are exogenous to the model and its choice can be considered arbitrary. Alternatively, the authors identify different market volatility regimes in an endogenous way and use high- and low-volatility regimes as proxies for bull and bear markets.

Extant studies applying dummy variables use different criteria to identify market states. For instance, Nofsinger and Varma (2014) and Silva and Cortez (2016) consider the peaks and troughs of the market index to identify periods of crisis and non-crisis. Carvalho and Areal (2016) and Leite *et al.* (2018) apply the procedure of Pagan and Soussonov (2003) to define periods of bull and bear market states. Besides distinguishing bull and bear markets on the basis of a market index, Henke (2016) also considers recessions and expansion periods according to the NBER business cycles and the Business Cycle Dating Committee for the Euro Area of the Centre for Economic Policy Research (CEPR). However, the definition of market states on the basis of the peaks and troughs of a market index or economic cycles defined by external institutions requires exogenous information, which is lagged relative to market data, i.e., they only become available "after the fact" (Kosowski, 2006). In response to some of these concerns, Kosowski (2006) and Areal *et al.* (2013) use the Markov-Switching model introduced by Hamilton (1989) to define market states in an endogenous way.

This dissertation is concerned with the issue of whether differences in ESG ratings are associated with differences in credit risks. Since company defaults and credit rating downgrades normally occur during economic recessions, we use the Business Cycle Dating Committee for the Euro Area of CEPR to define economic recessions as in Henke (2016). Although the use of the NBER or the CEPR recession and expansion periods also presents some drawbacks, there are several studies applying this criterion besides Henke (2016) (e.g., Moskowitz, 2010; Wang, 2010; Nofsinger and Varma, 2014). In addition, Kosowski (2006) concludes that the Markov Switching model confirms the findings based on NBER recession dates, and Carvalho and Areal (2016) state that the results based on NBER are consistent with those obtained with the Pagan and Sossounov (2003) procedure. Therefore, it seems that the use of CEPR recession and expansion periods is a simple and adequate criterion.

The incorporation of a dummy variable into Jensen's (1968) measure leads to the following regression:

$$r_{p,t} = \alpha_p + \alpha_{rec,p}D_t + \beta_{1p}r_{m,t} + \beta_{1rec,p}r_{m,t}D_t + \varepsilon_{p,t} \tag{12}$$

where D_t is a dummy variable which assumes the value of 0 in expansion periods and 1 in recession periods. The incorporation of dummy variables into the multi-factor model mentioned above leads to the following regression:

$$
\begin{aligned}
r_{p,t} = {} & \alpha_p + \alpha_{rec,p}D_t + \beta_{1p}Bond_t + \beta_{1rec,p}Bond_tD_t + \beta_{2p}Default_t \\
& + \beta_{2rec,p}Default_tD_t + \beta_{3p}Option_t + \beta_{3rec,p}Option_tD_t \\
& + \beta_{4p}Equity_t + \beta_{4rec,p}Equity_tD_t + \varepsilon_{p,t}
\end{aligned}
\tag{13}
$$

3.4. THE ESTIMATION PROCESS OF THE MODELS

The models are estimated by an Ordinary Least Squares (OLS) regression. There are several assumptions required to the Classical Linear Regression Model (CLRM) so that the OLS estimation technique has adequate properties and hypothesis tests can be applied properly (Brooks, 2014). However, it is usual to find at least one of these assumptions violated when financial models are estimated.

One of the assumptions is the existence of constant variance of the errors, known as the assumption of homoscedasticity. If the errors do not have a constant variance, they are said to be heteroscedastic. If OLS is still used in the presence of heteroscedasticity, the standard errors can be incorrect, leading to mistakes in statistical inference. In order to test for heteroscedasticity, we use the White (1980) test. It is a popular test because it is not based on the normality assumption (Brooks, 2014). In this case, the null hypothesis of homoscedasticity is rejected if the *p*-value is less than 5%.

Another assumption is that the covariance between the error terms over time is zero, i.e., it is assumed that the errors are uncorrelated with one another. If the errors are correlated, then there is autocorrelation. The consequences of ignoring autocorrelation are similar to those of ignoring heteroscedasticity. The coefficient estimates are still unbiased but the standard error estimates can be incorrect (Brooks, 2014). In order to test for autocorrelation, we use the Breusch (1978)–Godfrey (1978) / LM test. In this case, the null hypothesis of no correlation is rejected if the *p*-value is less than 5%.

According to Brooks (2014), the White (1980) variance–covariance matrix of the coefficients is appropriate when the residuals of the regression are heteroscedastic but serially uncorrelated. In the presence of both heteroscedasticity and autocorrelation, the Newey and West (1987) method will be used. The number of lags of the residuals to use is determined with the rule of thumb $\sqrt[4]{N}$, where N is the number of observations (Baum, 2006).

4. DATA

This chapter describes the data used in this dissertation as well as the respective sources. It starts by presenting the Asset4 ESG database, which provides information on companies' social scores, followed by the sample description as well as information on all variables used in the performance evaluation models.

4.1. DATA ON SOCIAL SCORES

Thomson Reuters ESG scores is an enhancement and replacement to the existing ASSET4 ratings. This database provides information on company social scores since 2002 and currently covers over 6.000 public companies across more than 400 ESG metrics. The Thomson Reuters ESG score is divided into 3 pillars and combined into 10 categories. All ESG data and scores are updated monthly. To interpret how companies are performing relative to their peers and where a company's ESG weaknesses and strengths lie, the score is available in both percentages and letter grades from D- to A+.[10] However, this database has not been released yet. Therefore, to measure the social responsibility level of a company, this study will use the ASSET4 ESG database. This database currently covers over 5.000 public companies and another major difference between the Thomson Reuters ESG and ASSET4 ESG is that the overall score in the latter case is divided into 4 pillars and combined into 18 categories.

ASSET4 ESG data is the choice of several recent studies on SRI performance (e.g., Halbritter and Dorfleitner, 2015; Stellner *et al.*, 2015; Gonenc and Scholtens, 2017). Gonenc and Scholtens (2017) point out some of the advantages of this source of social data, namely the consistency in the reporting since, for example, MSCI presents a structural break in the series in 2009. Further, only few situations of matching error should occur because Thomson

---

[10] For more detailed information about the Thomson Reuters ESG scores and methodology see https://financial.thomsonreuters.com/en/products/data-analytics/company-data/esg-research-data.html

Reuters also provides financial information about the companies (Gonenc and Scholtens, 2017).

Asset 4 ESG provides the equally-weighted ESG rating of a company considering in four pillars: economic, environmental, social and corporate governance pillars. According to Asset4, the economic pillar "represents a company's financial health and measures the sustainable growth and the return on investment". The environmental pillar "reflects how well a company uses best management practices to avoid environmental risks and capitalize on environmental issues". It includes several categories: *emission reduction, product innovation and resource reduction.* The social pillar "measures a company's capacity to generate trust and loyalty with its workforce, customers and society, and it is a reflection of the company's reputation". It includes *product responsibility, community, human rights, diversity, employment quality, health and safety, training and development* categories*.* The corporate governance pillar "measures a company's systems and processes, which ensures that its board members and executives act in the best interests of its long-term shareholders". It includes *board functions, board structure, compensation policy, shareholder rights, vision and strategy.*[11]

The environmental, social and corporate governance pillars represents companies' extra-financial health. Considering that the purpose of this dissertation is to form portfolios based on companies' extra-financial health, we combine the ESG scores into an equal weighted average of the individual scores (environment, social, and governance scores), as in Auer (2016).

4.2. SAMPLE

This study focuses on Eurozone firms covered by ASSET4 ESG between 2003 and 2016. As far as we are aware of, Hoepner and Nilsson (2017) is the only paper exploring the performance of fixed-income portfolios formed on the basis of ESG criteria. Yet, this paper addresses the US market alone. Therefore, we intend to explore this issue using a sample of

---

[11] Description of Thomson Reuters Asset4 database.

Euro-denominated bonds issued from Eurozone companies. The choice of the evaluation period was motivated by ASSET4 data availability.

Considering the Eurozone constituents identified from the ASSET4 database, Thomson Reuters DataStream was used to collect the financial information on corporate bonds issued by these firms. It was necessary to manually search for related securities in Thomson Reuters Datastream for each company, in order to include *active* and *dead* securities. Since total return data is preferred to using only price indices (because it takes into account capital gains and interest payments), this study used the end of month total return index from this database. As a consequence, only the last full month before the maturity date of each bond was considered. Even though a series of daily prices contains useful information that may be missing in a series of monthly prices, a set of monthly returns is closest to the normal distribution.[12] We also consider non-surviving firms and firms that are no longer public companies (identified from Thomson Reuters DataStream) in order to avoid *survivorship bias*. After collecting this information, we verified if the ASSET4 database covers these companies.

From the data retrieved, we excluded financial institutions or banks because these companies issue more securities than other industries and their inclusion would dominate the sample (Oikonomou *et al*., 2014; Hoepner and Nilsson, 2017). Ge and Liu (2015) also exclude financial firms because of different regulations and different debt financing characteristics. According to the ICB, the financial industry includes financial services, banks, insurance and real estate companies. As in Hoepner and Nilsson (2017) and Bauer and Hann (2010), we also excluded bonds with nonstandard characteristics that might result in their price being different (Elton *et al.*, 2001). These include "all floating rate notes, index-linked bonds, convertible bonds, exchangeable bonds, hybrids, preferred bonds, perpetual bonds, private placements, sinking fund provisions, bonds with embedded options or warrants, and bonds with any other nonstandard characteristic" (Oikonomou *et al.*, 2014, p. 60). Additionally, we excluded issues with asset-backed features because they represent the creditworthiness of the collateral instead of representing the creditworthiness of the issuer (Campbell and Taksler, 2003). Zero-coupon bonds were also excluded because they have

---

[12] There are three important properties that are observed in almost all sets of daily returns. First, the distribution of returns is not normal. Second, there is almost no correlation between returns for different days. Third, the correlations between the magnitudes of returns on nearby days are positive and statistically significant (Taylor, 2005). The three major stylized facts of returns are observed across time as well as across markets and, indeed, Mitchell *et al.* (2002) document these properties for the London fixed-income market (Taylor, 2005).

some characteristics of stocks (Bessembinder *et al.*, 2009). In addition, we did not consider issues from subsidiaries (including special purpose vehicles) when the company is the parent company. Finally, we included issues from the respective market as well as from international markets as long as the bond is Euro-denominated.

A company was eliminated from the sample if nothing is displayed under related securities from Thomson Reuters DataStream, either because the non-financial company did not issue any fixed-income securities without nonstandard characteristics, or because the bonds are not covered by the database.

The final sample includes 935 bonds issued by 189 companies.[13] The number of bonds and companies in each country and industry according to the ICB is presented in Appendix A. Table 1 provides some basic descriptive statistics of the individual ESG scores as well as of the combined ESG scores for the 189 Eurozone companies between 2003 and 2016. The mean ESG scores of all companies from Asset4 is likely to be around 50 because of their scoring method (Halbritter and Dorfleitner, 2015). On average, this sample has a mean environmental and social scores of 81.15 and 82.12, respectively. Therefore, it is possible to say that Eurozone companies are above average regarding these two pillars. From table 1, it is also possible to observe that these distributions are negatively skewed. This means that there are more high-rated companies than low-rated companies. Turning to the characteristics of the corporate governance score, Eurozone companies are in line with the average since the mean is 56.23. This also implies that Eurozone companies show worse scores in corporate governance pillar than they do in environmental and social pillars. Since the excess kurtosis is negative and the skewness is close to zero, it is possible to conclude that most companies are rated close to the middle-of-scale value of 50. The histogram of each ESG scores is presented in Appendix B.

---

[13] This final sample includes 37 bonds and 15 companies that are non-surviving firms or firms that are no longer public companies.

Table 1. Descriptive statistics of ESG scores

This table reports descriptive statistics of individual and combined ESG scores (equal weighted average of the three individual scores). The sample includes 189 Eurozone companies from the ASSET4 database between 2003 and 2016.

	Environmental Score	Social Score	Governance Score	Combined ESG Score
Mean	81.149	82.119	56.259	73.176
Std. Dev.	19.117	17.772	21.752	16.427
Median	89.434	88.375	60.075	77.836
Minimum	11.275	8.790	7.558	13.308
Maximum	96.423	96.910	94.878	94.631
Skewness	-2.097	-2.160	-0.356	-1.626
Kurtosis	6.783	7.704	2.145	5.553
Obs	189	189	189	189

Table 2 provides some basic descriptive statistics of the individual ESG scores as well as of the combined ESG scores for the high- and low-rated portfolios, comprising bonds from the 50% of all companies with the highest and lowest rating, respectively. The scores of each bond are averaged over time. The high-rated portfolios have a mean between 75.66 and 93.30. The low-rated portfolios have a mean between 46.21 and 80.72.[14]

---

[14] The descriptive statistics for ESG portfolios with alternative cut-offs are presented in Appendix C.

Table 2. Descriptive statistics of ESG scores on portfolios (positive screening strategy)

This table reports descriptive statistics on the individual and combined ESG scores of the high- and low-rated portfolios (positive screening strategy). The high-rated (low-rated) portfolios consists of bonds from the 50% of all companies with the highest (lowest) rating. The sample includes 935 bonds issued by 189 Eurozone companies from the ASSET4 database between 2003 and 2016.

	Environmental Score		Social Score		Governance Score		Combined ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	93.301	80.364	94.166	80.716	75.655	46.209	85.604	69.986
Std. Dev.	1.202	16.628	1.722	14.627	9.034	13.519	3.893	12.762
Median	93.422	87.148	94.564	85.928	75.283	50.218	85.489	74.339
Minimum	90.058	11.066	88.150	8.790	37.140	7.340	71.303	13.308
Maximum	97.164	92.623	97.320	93.900	95.910	69.440	94.637	82.887
Skewness	0.027	-2.394	-0.522	-2.617	-0.302	-0.870	-0.161	-2.254
Kurtosis	3.265	8.339	2.462	10.867	3.338	2.885	3.094	8.733
Obs	663	613	719	633	701	634	700	583

Table 3 provides some basic descriptive statistics on the credit ratings of the high- and low-rated portfolios, comprising 50% of bonds with the highest and lowest ratings for 616 of the 935 bonds included in the sample.[15] Following Oikonomou *et al.* (2014), we use the S&P Historical Bond Rating datatype from Thomson Reuters DataStream. The ordered ranking scale used ranges between one (for D rated bonds) and seven (for AAA rated bonds), as in Ashbaugh-Skaife *et al.* (2006).[16] The ranking scale is presented in appendix D. The credit ratings of each bond are averaged since they can change over time. The average bond observation has a score close to 4, representing the lowest tier of investment grade bonds (from BBB− to BBB+). It is worth mentioning that with the exception of the social score, all high-rated portfolios present a lower mean than low-rated portfolios. Indeed, it seems that high-rated portfolios are more exposed to speculative grade bonds (credit ratings lower than BBB-) than low-rated portfolios. The weights of investment and speculative grade bonds in each portfolio are presented in appendix E. As mentioned previously, since low credit rated bonds present high yields, issuers of speculative grade bonds can benefit the most in absolute terms of the reductions in the cost of debt that may arise from engagement in CSR practices (Oikonomou *et al.*, 2014). It seems that these issuers have a financial incentive to improve CSR practices.

---

[15] The remaining bonds are not rated by S&P.
[16] This approach is similar to the one used by Oikonomou *et al.* (2014).

Table 3. Descriptive statistics of credit ratings on portfolios (positive screening strategy)

This table provides some basic descriptive statistics of the credit ratings for 616 of the 935 bonds included in this sample. These ratings reflect S&P's assessment. The high-rated (low-rated) portfolios consists of bonds from the 50% of all companies with the highest (lowest) rating (positive screening strategy). The ordered ranking scale ranges between one for D (lowest rating) and seven for AAA (highest rating). The credit ratings of each bond are averaged.

	Environmental Score		Social Score		Governance Score		Combined ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	4.281	4.442	4.338	4.288	4.352	4.374	4.316	4.375
Std. Dev.	0.778	0.717	0.731	0.800	0.747	0.762	0.721	0.755
Median	4.000	4.438	4.120	4.000	4.176	4.091	4.000	4.008
Minimum	1.333	2.000	1.000	1.389	1.231	1.333	1.459	1.333
Maximum	6.000	6.000	6.000	6.017	6.000	6.039	6.000	6.036
Skewness	-0.560	-0.254	-0.701	-0.151	-0.346	-0.612	-0.185	-0.594
Kurtosis	3.813	2.826	4.850	3.197	3.347	3.931	3.389	4.073
Obs	483	385	511	399	490	405	493	358

Table 4 provides some basic descriptive statistics on the returns of the high- and low-rated portfolios, comprising 50% of bonds with the highest and lowest ratings, respectively, as well as of the long-short portfolio, which is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted as in Kempf and Osthoff (2007) and Hoepner and Nilsson (2017) and rebalanced monthly. However, as mentioned in chapter 3, we also intend to test equally-weighted portfolios to examine whether the results are sensitive to the portfolio weighting scheme. Besides the 50% cut-off, we additionally perform the skewness and kurtosis normality test[17] for alternative cut-off portfolios of 25% and 10%.[18] The results support the hypothesis that the 50% cut-off portfolio returns are normally distributed. We do not reject the null hypothesis of normality (at the 5% level) only for the high- and low-rated portfolios (with a 50% cut-off). There is more evidence of non-normal portfolios for alternative cut-offs. Anyhow, following the argument of Adcock *et al.* (2012), the non-normality of portfolio returns supports the use of conditional models rather than unconditional models.

Table 4 shows similar patterns for the returns of portfolios formed on the basis of alternative metrics of CSR. The high- and low-rated portfolios present positive and statistically significant monthly returns. Although high-rated portfolios show higher monthly returns than the low-rated portfolios, the difference is not statistically significant. Thereby, it is possible to conclude that the mean return does not differ significantly between the high- and low-rated portfolios. These findings are consistent regardless the ESG metric used to form the portfolios. The results (shown in appendix F) are similar for alternative cut-off portfolios of 25% and 10%.

---

[17] Using the Stata software, we perform the skewness and kurtosis normality test with a statistic that is approximately a χ^2 distribution with 2 degrees of freedom under the null hypothesis of normality. This test is similar to the Jarque–Bera (1987) test of normality. Both tests are calculated from the sample skewness and kurtosis, though the Jarque-Bera test is based on asymptotic standard errors with no corrections for sample size. In turn, the skewness and kurtosis normality test used performs two adjustments for sample size: that of Royston (1991) and that of D'Agostino, *et al.* (1990) (Stata Press, 2017).
[18] The descriptive statistics for portfolios formed with the 25% and 10% cut-off are presented in Appendix F.

Table 4. Descriptive statistics of portfolios (positive screening strategy)

This table reports descriptive statistics on the monthly returns of the high- and low-rated portfolios (50% cut-off) as well as the long-short portfolios for each individual and combined ESG scores between 2003 and 2016. The high-rated (low-rated) portfolios consists of bonds from the 50% of all companies with the highest (lowest) rating and the long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio (positive screening strategy). The portfolios are value-weighted and rebalanced monthly. The Adj. $\mathcal{X}^2$ is a statistic that is approximately a $\mathcal{X}^2$ distribution with 2 degrees of freedom under the null of normality. The result "." should be interpreted as an absurdly large number and, hence, the data are most certainly not normal. ***, **and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Mean	Std. Dev.	Median	Minimum	Maximum	Skewness	Kurtosis	Adj. $\mathcal{X}^2$	*p*-value
Environmental Score									
High-rated	0.438%***	0.898%	0.517%	-2.356%	2.508%	-0.165	2.887	0.830	0.660
Low-rated	0.428%***	0.883%	0.517%	-1.813%	2.394%	-0.202	2.598	2.480	0.289
Long-short	0.010%	0.358%	-0.019%	-1.041%	2.779%	2.699	24.053	.	0.000
Social Score									
High-rated	0.445%***	0.892%	0.528%	-2.203%	2.470%	-0.155	2.845	0.770	0.681
Low-rated	0.444%***	0.872%	0.547%	-1.894%	2.313%	-0.199	2.631	2.160	0.339
Long-short	0.001%	0.378%	-0.010%	-1.034%	3.158%	3.692	32.168	.	0.000
Governance Score									
High-rated	0.443%***	0.923%	0.538%	-2.400%	2.776%	-0.070	2.920	0.150	0.929
Low-rated	0.417%***	0.830%	0.554%	-1.765%	2.390%	-0.194	2.550	2.910	0.234
Long-short	0.026%	0.356%	0.018%	-1.532%	1.916%	0.571	9.777	31.760	0.000
Combined ESG Score									
High-rated	0.448%***	0.904%	0.559%	-2.270%	2.579%	-0.118	2.841	0.470	0.789
Low-rated	0.412%***	0.867%	0.547%	-2.228%	2.224%	-0.336	2.839	3.350	0.187
Long-short	0.036%	0.378%	0.011%	-1.027%	3.188%	3.428	31.018	.	0.000

4.3. OTHER DATA

Following Leite and Cortez (2018), the main set of benchmark indices are the iBoxx Total Return (TR) bond index family and the BofA Merrill Lynch Total Return (TR) bond index family. The BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index is used as the bond index.[19] The default spread is computed as the difference in returns between the BofA Merrill Lynch € High-Yield TR index[20] and the iBoxx € Sovereign TR index. The option variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. Finally, the stock market variable is measured by the excess returns of the FTSE AW Eurozone TR index. Excess returns for fund returns and benchmarks were computed using the 1-month Euribor. Data on indices was retrieved from Thomson Reuters Datastream.[21]

Table 5 presents some basic descriptive statistics on the monthly returns of the factors used in the performance evaluation models between 2003 and 2016. It is possible to observe that the risk-free rate (R_f) presents a lower mean and standard deviation than the market index over the period under analysis. As expected, the default spread and the equity factor present a positive mean and, hence, reflect default risk compensation and equity risk premium. The null hypothesis of normality is rejected at the 10% level for the bond index and at 1% for the remaining factors. The correlations between the factors are presented in Appendix G. They are relatively small, except for the equity factor and default spread (0.647).

---

[19] This index covers the eurozone market.

[20] The iBoxx € High Yield index does not cover the entire sample period.

[21] The exception is the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile, since   data on this index is retrieved from https://markets.ml.com/search-results?p_p_id=MercurySearch_WAR_Mercurysearchportlet&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_MercurySearch _WAR_Mercurysearchportlet__spage=%2Fportlet_action%2Fsearch%2Fgenxsearch.

Table 5. Descriptive statistics on the factors used in the performance evaluation models

This table reports descriptive statistics on the monthly returns of the bond market index, default factor – return spread between a high-yield bond index and a government bond index, option factor - return difference between a mortgage-backed securities index and a government bond index, equity factor - excess returns of the equity market index, and the risk-free rate (R_f). The Adj. $\mathcal{X}^2$ is a statistic that is approximately a $\mathcal{X}^2$ distribution with 2 degrees of freedom under the null of normality.

	Market	**Default**	**Option**	**Equity**	R_f
Mean	0.426%	0.407%	-0.004%	0.561%	0.146%
Std. Dev.	0.957%	3.389%	0.799%	4.826%	0.123%
Median	0.504%	0.387%	-0.003%	1.270%	0.170%
Minimum	-3.104%	-19.557%	-3.199%	-16.446%	-0.013%
Maximum	2.802%	13.600%	2.633%	15.987%	0.421%
Skewness	-0.375	-0.947	0.007	-0.493	0.392
Kurtosis	3.372	11.974	5.253	4.280	1.901
Adj. $\mathcal{X}^2$	5.310	44.460	11.030	11.540	36.010
p-value	0.070	0.000	0.004	0.003	0.000
Obs	168	168	168	168	168

Regarding the public information variables, the term spread variable is measured by the annualized yield spread between the European Monetary Union Benchmark 10 Years Datastream Government (iBoxx) and the annualized 3-month Euribor rate. To calculate the IRW variable, the past real wealth for the Euro-Area is estimated by an exponentially weighted average of past levels of the FTSE AW Eurozone index in EUR deflated by the Euro-Area Harmonised Index of Consumer Prices (HICP). Data on the public information variables to be used in the conditional models are also collected from Thomson Reuters Datastream.

Table 6 provides some basic descriptive statistics on the information variables used in the conditional models between 2002 and 2016. The null hypothesis of normality is not rejected at the 1% level only for the term spread and the transformed IRW. The correlation between the information variables is relatively small and is presented in Appendix G.

Table 6. Descriptive statistics on the information variables used in the conditional models

This table reports descriptive statistics on the monthly returns of the term spread (TS) - difference between the yield of a long-term bond and the yield of a short-term bond, and the inverse relative wealth (IRW) - ratio between the exponentially weighted average of past real wealth and current wealth, before and after the variable being transformed using the stochastic detrending and mean-zero procedure. The Adj. $\mathcal{X}^2$ is a statistic that is approximately a $\mathcal{X}^2$ distribution with 2 degrees of freedom under the null of normality.

	TS	TS (transformed)	IRW	IRW (transformed)
Mean	1.018%	0.000%	0.998	0.000
Std. Dev.	0.845%	0.531%	0.158	0.133
Median	1.035%	-0.085%	0.947	-0.004
Minimum	-1.265%	-1.147%	0.816	-0.384
Maximum	2.684%	2.007%	1.651	0.422
Skewness	-0.266	1.462	1.893	0.157
Kurtosis	2.580	6.270	6.758	4.451
Adj. $\mathcal{X}^2$	3.530	42.330	54.310	7.840
***p*-value**	0.171	0.000	0.000	0.020
Obs	168	168	168	168

Regarding the states of the economy, as mentioned before, we use the Business Cycle Dating Committee for the Euro Area of CEPR to define economic recessions, as in Henke (2016). For the sample period CEPR identifies 33 months of economic recessions from January 2008 until April 2009 and from July 2011 until January 2013[22] (The Business Cycle Dating Committee for the Euro Area of the Centre for Economic Policy Research, 2017).

---

[22] Regarding the first recession, the CEPR Euro Area Business Cycle Dating Committee declared that the month of the trough is April 2009 and the month of the peak is January 2008. Regarding the second recession, the committee declared that the trough of the recession that started after the 2011Q3 peak has been reached in 2013Q1.

5. EMPIRICAL RESULTS

This chapter starts by presenting the results on performance of value-weighted portfolios consisting of European bonds formed on the basis of a positive screening strategy. The performance of these portfolios, formed with a 50% cut-off for the individual and combined ESG scores, is evaluated with multifactor and and conditional models. Several robustness tests are also performed. Moreover, this chapter addresses the evolution of social and financial performance over time as well as portfolio performance during periods of crisis. This chapter proceeds with an analysis of the results of portfolios formed on the basis of a best-in-class strategy.

5.1. SRI BOND PORTFOLIOS FORMED BY USING A POSITIVE SCREENING STRATEGY

5.1.1 PERFORMANCE EVALUATION OF SRI BOND PORFOLIOS

We start by presenting the portfolio performance results obtained from the multi-factor model (equation 5).

Table 7 provides the regression results of the multi-factor model of high- and low-rated portfolios as well as long-short portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off. As expected, bond market risk has a significant impact (at the 1% level) on the excess returns of the high- and low-rated portfolios, regardless of the score used.[23] The results suggest a lower-than-average sensitivity to the market index, since beta is always lower than 1. While high-rated portfolios are not significantly exposed to the default factor, this factor has negative and a significant influence on the excess returns of low-rated porfolios. The results also show statistically significant differences in investment

---

[23] We also run the multi-factor model regression with a broader bond market index (the iBoxx € Overall) as the bond market benchmark. As expected, the explanatory power decreases for all portfolios. The regression results are presented in appendix H.

styles between high- and low rated portfolios. Although high- and low-rated portfolios present similar exposures to the bond market factor, high-rated portfolios are significantly more exposed to the default factor (except for portfolios formed on the social score). In turn, high-rated portfolios are significantly less exposed to the equity factor (except for the corporate governance score) and are significantly less exposed to the option factor in the case of portfolios formed on the social score.

Table 7 also shows that both high- and low-rated portfolios formed on the basis of the social score yield a positive and statistically significant abnormal return. Yet, the alphas of the long-short portfolios are not statistically significant for any of the portfolios, indicating that investors cannot obtain abnormal returns by going long in high-rated bonds and short in low-rated bonds.

In the fixed-income area, although Hoepner and Nilsson (2017) do not find statistically significant results for the aggregated high- and low-rated portfolios, they also conclude that it is not possible to earn abnormal returns by employing the long-short strategy. These initial findings are also in line with those of Kempf and Osthoff (2007) for equity portfolios, since they also do not find an outperformance of the long-short strategy by forming portfolios on the basis of a 50% cut-off. Halbritter and Dorfleitner (2015), also using ESG data of ASSET4 for equity portfolios, find that abnormal returns of high- and low-rated as well as of the long-short porfolios for each individual and combined ESG scores are statistically insignificant.

The initial findings of this study do not find any evidence that portfolios of bonds issued by companies with high ESG ratings (either at the individual or combined ESG scores) outperform porfolios of companies with low ESG ratings. However, the unconditional model used suffers from some important limitations, as mentioned previously, since it assumes constant alphas and betas over time. Hence, we next present the results obtained from the conditional performance evaluation model.

Table 7. The performance of portfolios formed on individual and combined ESG scores (positive screening strategy) - multifactor model

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the multi-factor model regressions (equation 5). *Bond* corresponds to the monthly excess returns of the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$
Env. Score						
High-rated	0.043	0.883***	0.000	0.007	0.003	0.902
	(1.62)	(27.78)	(0.00)	(0.15)	(0.33)	
Low-rated	0.047*	0.889***	-0.061***	0.079**	0.020***	0.867
	(1.83)	(30.58)	(-5.34)	(2.02)	(2.69)	
Long-short	-0.004	-0.006	0.061***	-0.072*	-0.017**	0.150
	(-0.15)	(-0.20)	(5.18)	(-1.80)	(-2.27)	
Soc. Score						
High-rated	0.059**	0.881***	-0.020	0.007	0.003	0.887
	(2.06)	(26.40)	(-1.09)	(0.14)	(0.40)	
Low-rated	0.058**	0.858***	-0.028**	0.092**	0.021***	0.855
	(2.17)	(28.82)	(-2.38)	(2.31)	(2.81)	
Long-short	0.001	0.023	0.008	-0.086**	-0.018**	0.041
	(0.04)	(0.70)	(0.55)	(-2.32)	(-2.12)	
Gov. Score						
High-rated	0.043	0.899***	-0.011	0.005	0.011	0.885
	(1.47)	(26.30)	(-0.54)	(0.11)	(1.23)	
Low-rated	0.045*	0.860***	-0.047***	0.080**	0.007	0.908
	(1.87)	(31.36)	(-2.91)	(2.47)	(0.98)	
Long-short	-0.002	0.039	0.036**	-0.075	0.004	0.154
	(-0.09)	(1.29)	(2.48)	(-1.58)	(0.36)	
ESG Score						
High-rated	0.051*	0.894***	-0.005	0.015	0.004	0.903
	(1.83)	(27.44)	(-0.25)	(0.32)	(0.47)	
Low-rated	0.038	0.861***	-0.063**	0.078*	0.022***	0.844
	(1.39)	(27.85)	(-5.13)	(1.88)	(2.77)	
Long-short	-0.004	0.033	0.058***	-0.063	-0.018**	0.137
	(0.45)	(1.04)	(4.60)	(-1.49)	(-2.24)	

The conditional model of Ferson and Schadt (1996) allows betas to be time-varying but the alpha remains constant. Christopherson *et al.* (1998) extend the model to a full conditional specification by allowing for time-varying conditional betas and alphas. According to Ferson *et al.* (2008), when time-varying alphas are not included in the model and only time-varying betas are considered, estimates of betas will be biased. Therefore, we choose to apply the full conditional specification of the multi-factor model (equation 10).

Table 8 provides the regression results of the conditional multi-factor model for high- and low-rated portfolios as well as the long-short portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off.

In comparison with the results obtained with the unconditional version of this model, the incorporation of the lagged information variables slightly increases the explanatory power of all portfolios. The increase of the explanatory power is consistent with the results of most empirical studies using conditional performance measures on the equity market (e.g., Ferson and Schadt, 1996; Bauer *et al.*, 2005 and Bauer *et al.*, 2006). Regarding the fixed-income area, Leite and Cortez (2018) also find that the conditional model provides higher adjusted R^2 for SRI bond funds. Although Silva *et al.* (2003) find that the adjusted R^2 of the portfolios of funds remain similar or decreases slightly, they reject the hypothesis that the coefficients for the additional variables are jointly equal to zero for several funds.

The results of the Wald test allow us to reject the hypothesis of the conditional alphas being equal to zero for the high-rated portfolios formed on the environmental and social scores (at the 5% and 1% level, respectively). For the low-rated portfolios, the Wald test also suggests rejecting the hypothesis of the conditional alphas being equal to zero (at the 5% level) for the portfolio based on the government score. In these cases, it is possible to conclude that alphas are time-varying. Regarding conditional betas (which are the interaction terms between the benchmark index and the lagged conditioning variables), the results of the Wald test suggest rejecting the null hypothesis that these coefficients are equal to zero for the low-rated portfolio formed on the governance score. Finally, the Wald test suggests rejecting the hypothesis of the conditional alphas and betas being jointly equal to zero for the high-rated portfolios formed on the environmental and social score. For the low-rated portfolios, the results of the Wald test suggest rejecting the null hypothesis that these

coefficients are jointly equal to zero (at the 1% level) only for the portfolio based on the governance score. In general, the results of the Wald test support the use of a conditional model with time-varying alphas and betas.

As previously documented in the case of table 7, bond market risk has a significant impact (at the 1% level) on the excess returns of the high- and low-rated portfolios, regardless of the score used. Also, high-rated portfolios are not significantly exposed to the default factor, whereas low-rated portfolios show a negative and statistically significant coefficient on this factor (except for the one formed on the basis of the social score). The results also show that there are statistically significant differences in investment styles between high- and low rated portfolios. Although high- and low-rated portfolios present similar exposures to the bond market, the results show that high-rated portfolios are significantly more exposed to the default factor for portfolios formed on the basis of environmental and ESG scores. As mentioned previously, since low credit rated bonds present high yields, issuers of speculative grade bonds can benefit the most in absolute terms from reductions in the cost of debt that may result from CSR practices (Oikonomou *et al.*, 2014). In addition, Stellner *et al.* (2015), based on a sample of Eurozone corporate bond market, find that companies with superior CSR benefit from better ratings and lower spreads in countries with above average ESG scores. It does sound like these issuers have a financial incentive to improve on CSR practices.

Regarding the information variables, the coefficients of the term spread and the IRW are statistically significant for the high-rated portfolios (with the exception of the IRW of the portfolio formed on the corporate governance score, which is statistically significant only at the 10% level). As expected, the coefficients of the term spread are positive because expected returns on corporate bonds have a positive relationship with the variation in maturity premiums (differences between the expected returns on long- and short-term bonds). This result is consistent to what is usually observed in the US market but contrasts with the results of Silva *et al.* (2003) for Europe since the authors find a negative coefficient for this information variable. As also expected, the coefficients of the IRW variable are positive, since risk aversion measured by the IRW is positively correlated with expected bond returns. In contrast, Silva *et al.* (2003) find a significant negative relation between the conditional beta and IRW, although they find a positive coefficient in their analysis of bond return

predictability. The January dummy variable is positive and statistically significant (at the 5% level) for low-rated portfolios formed on the governance and ESG scores.

The results show that high-rated portfolios formed on the basis of the social and ESG scores yield a positive and statistically significant abnormal return (at the 5% level). The low-rated portfolios do not show positive alphas. All high-rated portfolios show higher alphas than the low-rated portfolios. Yet, the differences are not statistically significant. Hence, the results suggest that investors cannot obtain abnormal returns by going long in high-rated stocks and short in low-rated stocks.

Table 8. The performance of portfolios formed on individual and combined ESG scores (positive screening strategy) - conditional multifactor model

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). *Bond* corresponds to the monthly excess returns of the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

	α	α_{TS}	α_{IRW}	α_{JD}	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$\beta_{Bond*TS}$	β_{Def*TS}	β_{Opt*TS}	β_{Eq*TS}
Env. Score												
High-rated	0.054*	0.002**	0.009**	0.001	0.869***	0.009	0.045	0.004	-0.132**	-0.020	0.055	-0.001
	(1.97)	(2.18)	(2.51)	(1.40)	(26.59)	(0.79)	(1.50)	(0.64)	(-2.17)	(-1.00)	(0.75)	(-0.05)
Low-rated	0.027	0.000	0.004	0.000	0.903***	-0.044***	0.072*	0.021**	-0.080	0.019	-0.053	-0.019
	(0.89)	(0.52)	(1.27)	(-0.06)	(28.72)	(-2.63)	(1.66)	(2.37)	(-1.17)	(0.92)	(-0.69)	(-1.39)
Long-short	0.027	0.001	0.005*	0.001	-0.034	0.053***	-0.027	-0.017*	-0.052	-0.038*	0.107	0.019
	(0.87)	(1.56)	(1.71)	(1.10)	(-1.06)	(3.11)	(-0.60)	(-1.87)	(-0.74)	(-1.83)	(1.38)	(1.32)
Soc. Score												
High-rated	0.074**	0.002**	0.011***	0.000	0.877***	-0.001	0.043	0.005	-0.162***	-0.031	0.070	0.002
	(2.47)	(2.51)	(3.21)	(0.25)	(29.63)	(-0.07)	(1.61)	(0.82)	(-2.75)	(-1.62)	(0.93)	(0.14)
Low-rated	0.027	0.000	0.002	0.001	0.861***	-0.013	0.071	0.017*	0.034	0.022	-0.047	-0.014
	(0.87)	(-0.14)	(0.62)	(0.99)	(26.57)	(-0.75)	(1.59)	(1.89)	(0.48)	(1.03)	(-0.61)	(-1.01)
Long-short	0.047	0.002**	0.009***	-0.001	0.017	0.012	-0.028	-0.013	-0.196**	-0.053**	0.117	0.016
	(1.36)	(2.12)	(2.89)	(-0.67)	(0.47)	(0.64)	(-0.58)	(-1.27)	(-2.54)	(-2.29)	(1.37)	(1.02)

Table 8. Continued

	$\beta_{Bond*IRW}$	$\beta_{Def*IRW}$	$\beta_{Opt*IRW}$	β_{Eq*IRW}	$\beta_{Bond*JD}$	β_{Def*JD}	β_{Opt*JD}	β_{Eq*JD}	$Adj.R^2$	W_1	W_2	W_3
Env. Score												
High-rated	-0.413	-0.103	0.142	0.039	0.035	-0.121	0.029	0.050	0.907	0.026	0.465	0.007
	(-1.59)	(-1.12)	(0.50)	(0.80)	(0.30)	(-0.79)	(0.25)	(0.82)				
Low-rated	-0.709***	-0.092	-0.120	-0.004	-0.141	-0.157	-0.100	0.072*	0.871	0.648	0.090	0.175
	(-2.86)	(-1.05)	(-0.35)	(-0.07)	(-0.68)	(-1.48)	(-0.56)	(1.67)				
Long-short	0.296	-0.011	0.263	0.043	0.176	0.036	0.129	-0.022	0.183	0.199	0.254	0.137
	(1.16)	(-0.12)	(0.73)	(0.74)	(0.83)	(0.33)	(0.70)	(-0.49)				
Soc. Score												
High-rated	-0.810***	-0.113	0.148	0.008	-0.052	-0.129	-0.062	0.065	0.898	0.004	0.051	0.023
	(-3.46)	(-1.28)	(0.56)	(0.15)	(-0.43)	(-0.80)	(-0.48)	(1.07)				
Low-rated	-0.277	-0.103	-0.127	0.061	0.075	-0.172	0.155	0.062	0.858	0.672	0.245	0.253
	(-1.08)	(-1.15)	(-0.35)	(1.04)	(0.35)	(-1.57)	(0.84)	(1.39)				
Long-short	-0.533*	-0.009	0.275	-0.054	-0.127	0.042	-0.217	0.003	0.113	0.026	0.037	0.030
	(-1.91)	(-0.09)	(0.70)	(-0.84)	(-0.55)	(0.35)	(-1.07)	(0.06)				

Table 8. Continued

	α	α_{TS}	α_{IRW}	α_{JD}	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$\beta_{Bond*TS}$	β_{Def*TS}	β_{Opt*TS}	β_{Eq*TS}
Gov. Score												
High-rated	0.047	0.002**	0.007*	0.000	0.891***	0.005	0.026	0.011	-0.087	-0.020	0.027	-0.001
	(1.57)	(2.38)	(1.71)	(-0.35)	(28.79)	(0.28)	(0.68)	(1.16)	(-1.44)	(-0.96)	(0.31)	(-0.08)
Low-rated	0.032	0.000	0.005*	0.002**	0.874***	-0.036**	0.094***	0.009	-0.174***	0.021	-0.044	-0.019*
	(1.42)	(0.59)	(1.81)	(2.55)	(37.21)	(-2.37)	(2.89)	(1.17)	(-3.28)	(1.65)	(-0.85)	(-1.73)
Long-short	0.015	0.001*	0.002	-0.003**	0.017	0.041*	-0.067	0.002	0.087**	-0.041**	0.071	0.017
	(0.57)	(1.76)	(0.74)	(-2.02)	(0.63)	(1.84)	(-1.37)	(0.17)	(1.99)	(-2.18)	(0.78)	(1.39)
ESG Score												
High-rated	0.059**	0.002**	0.009**	0.000	0.885***	0.012	0.043	0.003	-0.100*	-0.025	0.035	0.004
	(2.08)	(2.49)	(2.22)	(-0.23)	(28.97)	(0.75)	(1.09)	(0.42)	(-1.73)	(-1.43)	(0.45)	(0.32)
Low-rated	0.018	0.000	0.003	0.002**	0.872***	-0.056***	0.081*	0.026***	-0.133*	0.033	-0.042	-0.029**
	(0.58)	(0.37)	(0.96)	(2.03)	(26.47)	(-3.21)	(1.77)	(2.74)	(-1.86)	(1.55)	(-0.52)	(-2.04)
Long-short	0.040	0.001	0.006*	-0.003**	0.013	0.068***	-0.037	-0.022**	0.033	-0.058***	0.076	0.033**
	(1.23)	(1.62)	(1.88)	(-2.20)	(0.39)	(3.79)	(-0.80)	(-2.31)	(0.45)	(-2.64)	(0.94)	(2.24)

Table 8. Continued

	$\beta_{Bond*IRW}$	$\beta_{Def*IRW}$	$\beta_{Opt*IRW}$	β_{Eq*IRW}	$\beta_{Bond*JD}$	β_{Def*JD}	β_{Opt*JD}	β_{Eq*JD}	$Adj.R^2$	W_1	W_2	W_3
Gov. Score												
High-rated	-0.345	-0.153	0.149	0.027	-0.212	-0.125	-0.171	0.073	0.886	0.115	0.512	0.225
	(-0.95)	(-1.41)	(0.40)	(0.39)	(-1.29)	(-0.69)	(-0.96)	(1.07)				
Low-rated	-0.860***	-0.022	-0.121	-0.010	0.170	-0.166	0.156	0.049	0.925	0.021	0.004	0.000
	(-3.49)	(-0.35)	(-0.47)	(-0.25)	(0.97)	(-1.22)	(1.04)	(0.85)				
Long-short	0.516**	-0.131*	0.271	0.038	-0.382	0.040	-0.327	0.024	0.217	0.076	0.001	0.000
	(2.50)	(-1.74)	(0.96)	(0.51)	(-1.51)	(0.52)	(-1.63)	(0.62)				
ESG Score												
High-rated	-0.404	-0.131	0.173	0.017	-0.135	-0.138	-0.132	0.080	0.907	0.068	0.303	0.108
	(-1.16)	(-1.36)	(0.53)	(0.28)	(-0.98)	(-0.88)	(-0.88)	(1.35)				
Low-rated	-0.812***	-0.042	-0.201	0.021	0.115	-0.162	0.152	0.036	0.853	0.186	0.051	0.055
	(-3.13)	(-0.46)	(-0.55)	(0.35)	(0.54)	(-1.45)	(0.81)	(0.78)				
Long-short	0.41	-0.089	0.374	-0.004	-0.250	0.024	-0.284	0.044	0.191	0.029	0.186	0.051
	(1.53)	(-0.94)	(1.00)	(-0.07)	(-1.13)	(0.21)	(-1.47)	(0.95)				

Compared with the unconditional model, the results of the conditional model show a slight tendency for higher alphas of high-rated portfolios and a tendency for lower alphas on low-rated portfolios. In addition, there is a slight tendency for higher alphas on the long-short portfolios. Yet, the alphas of the long-short portfolios are still not statistically significant, suggesting that the performance of portfolios of bonds issued by companies with high ESG ratings (either at the individual or combined ESG scores) is not statistically different from that of porfolios of companies with low ESG ratings. Table 9 summarizes and compares the alphas obtained with the unconditional and conditional models.

Table 9. Comparison of alphas between the unconditional and conditional models (positive screening strategy)

This table summarizes the comparison of alphas (expressed in percentage) between the unconditional and conditional models. ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis.

	Unconditional model			Conditional model		
	High-rated	**Low-rated**	**Long-short**	**High-rated**	**Low-rated**	**Long-short**
Env. Score	0.043	0.047*	-0.004	0.054*	0.027	0.027
	(1.62)	(1.83)	(-0.15)	(1.97)	(0.89)	(0.87)
Soc. Score	0.059**	0.058**	0.001	0.074**	0.027	0.047
	(2.06)	(2.17)	(0.04)	(2.47)	(0.87)	(1.36)
Gov. Score	0.043	0.045*	-0.002	0.047	0.032	0.015
	(1.47)	(1.87)	(-0.09)	(1.57)	(1.42)	(0.57)
ESG Score	0.051*	0.038	-0.004	0.059**	0.018	0.040
	(1.83)	(1.39)	(0.45)	(2.08)	(0.58)	(1.23)

5.1.2   ROBUSTNESS TESTS

In this section, several robustness tests are performed in order to test the sensitivity of the results to alternative cut-off-portfolios, the exclusion of outliers, and an alternative portfolio weighting scheme. We choose to continue with the full conditional specification of the multi-factor model (equation 10) since there is a slight increase of the explanatory power in comparison with the results obtained with the unconditional version of this model and since we find some evidence of time varying betas and alphas.

5.1.2.1   Performance of alternative cut-off portfolios

Extant empirical evidence shows that the the profitability of the long-short strategy can depend on the cut-off chosen to form portfolios. For instance, Schröder (2014) points out that studies on SRI funds and indices do not find an outperformance of SRI strategies in most cases because they include not only the few companies with a very good CSR rating but also a high number of companies with a mediocre rating. Furthermore, Kempf and Osthoff (2007) only find a significant outperformance for portfolios formed on the basis of extreme social ratings (e. g., the 10% companies with the best rating). Using the top 50% of social ratings, the results cease to be statistically significant (Kempf and Osthoff, 2007). Hence, we also estimate the results obtained from portfolios formed on the basis of alternative 25% and 10% cut-offs.

As mentioned previously, there is more evidence of non-normal distribution of returns for portfolios based on these alternative cut-offs. Anyhow, following the argument of Adcock *et al.* (2012), the non-normality of portfolio returns supports the use of conditional models rather than unconditional models. Therefore, we choose to apply the full conditional specification of the multi-factor model (equation 10).

Table 10 provides the regression results of the conditional multi-factor model of high- and low-rated portfolios as well as the long-short portfolios formed on the basis of individual and combined ESG scores with a 25% (Panel A) and a 10% (Panel B) cut-off. The low-rated

portfolio based on ESG scores with a 10% cut-off presents a considerably low adjusted R^2 (0.141).[24]

---

[24] This might reflect the low number of companies comprising the 10% worst porftolios in some months.

Table 10. The performance of portfolios formed on individual and combined ESG scores (positive screening strategy) - alternative cut-offs

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). *Bond* corresponds to the monthly excess returns of the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 25% and 10% of all companies with the highest (lowest) ratings for each individual and combined ESG scores (panel A and B, respectively). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

PANEL A	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.030	0.839***	0.000	0.086**	0.007	0.898	0.144	0.000	0.000
	(1.12)	(26.34)	(0.03)	(2.18)	(1.05)				
Low-rated	0.013	0.869***	-0.048*	0.039	0.027*	0.679	0.572	0.762	0.800
	(0.25)	(16.02)	(-1.68)	(0.52)	(1.77)				
Long-short	0.017	-0.030	0.049*	0.047	-0.020	0.034	0.080	0.475	0.326
	(0.33)	(-0.56)	(1.70)	(0.63)	(-1.28)				
Soc. Score									
High-rated	0.077*	0.900***	-0.017	0.048	0.008	0.874	0.012	0.137	0.051
	(1.82)	(27.74)	(-1.18)	(1.27)	(1.07)				
Low-rated	0.037	0.792***	-0.022	0.033	0.027*	0.706	0.781	0.696	0.774
	(0.76)	(15.83)	(-0.84)	(0.48)	(1.94)				
Long-short	0.041	0.107*	0.005	0.015	-0.019	0.151	0.041	0.000	0.000
	(0.58)	(1.93)	(0.19)	(0.25)	(-1.09)				
Gov. Score									
High-rated	0.009	0.888***	0.019	-0.021	0.007	0.858	0.588	0.808	0.448
	(0.27)	(26.05)	(0.97)	(-0.42)	(0.64)				
Low-rated	0.056*	0.846***	-0.055**	0.138***	0.015	0.852	0.202	0.302	0.130
	(1.73)	(23.74)	(-2.54)	(2.80)	(1.40)				
Long-short	-0.047	0.042	0.074***	-0.159***	-0.009	0.369	0.112	0.001	0.001
	(-1.25)	(1.08)	(3.54)	(-2.93)	(-0.79)				
ESG Score									
High-rated	0.058*	0.910***	0.022	0.002	0.003	0.870	0.137	0.431	0.144
	(1.70)	(26.08)	(1.07)	(0.04)	(0.30)				
Low-rated	0.049	0.725***	-0.058*	0.093	0.039**	0.497	0.720	0.734	0.834
	(0.77)	(11.04)	(-1.67)	(1.03)	(2.11)				
Long-short	0.009	0.185***	0.080**	-0.091	-0.036*	0.132	0.217	0.496	0.430
	(0.14)	(2.84)	(2.32)	(-1.01)	(-1.94)				

Table 10. Continued

PANEL B	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.041	0.783***	0.003	0.117**	0.009	0.816	0.139	0.000	0.000
	(1.35)	(18.77)	(0.12)	(2.15)	(0.96)				
Low-rated	0.011	0.765***	-0.088	0.013	0.059*	0.302	0.312	0.856	0.869
	(0.11)	(7.03)	(-1.53)	(0.09)	(1.91)				
Long-short	0.029	0.018	0.091	0.104	-0.049*	-0.004	0.084	0.485	0.415
	(0.29)	(0.17)	(1.64)	(0.72)	(-1.67)				
Soc. Score									
High-rated	0.088*	0.932***	-0.005	0.051	0.009	0.820	0.045	0.156	0.093
	(1.69)	(23.04)	(-0.27)	(1.09)	(0.86)				
Low-rated	0.065*	0.784***	-0.033	0.126**	0.013	0.686	0.116	0.284	0.022
	(1.76)	(13.06)	(-1.12)	(2.29)	(0.92)				
Long-short	0.024	0.148***	0.027	-0.075	-0.004	0.188	0.029	0.023	0.001
	(0.42)	(2.61)	(0.88)	(-1.03)	(-0.26)				
Gov. Score									
High-rated	-0.024	0.901***	0.015	-0.030	0.000	0.815	0.596	0.388	0.317
	(-0.52)	(19.94)	(0.61)	(-0.58)	(-0.02)				
Low-rated	0.051	0.674***	-0.002	0.168**	0.025	0.494	0.175	0.591	0.357
	(0.89)	(11.44)	(-0.08)	(2.06)	(1.50)				
Long-short	-0.074	0.227***	0.018	-0.198**	-0.025	0.151	0.437	0.264	0.252
	(-1.12)	(3.30)	(0.49)	(-2.08)	(-1.29)				
ESG Score									
High-rated	-0.009	0.915***	0.012	-0.062	0.006	0.845	0.509	0.227	0.156
	(-0.24)	(22.40)	(0.52)	(-1.03)	(0.55)				
Low-rated	0.017	0.712***	-0.142*	0.096	0.079**	0.141	0.579	0.980	0.982
	(0.12)	(5.07)	(-1.91)	(0.49)	(1.99)				
Long-short	-0.026	0.203	0.154**	-0.158	-0.073*	0.000	0.380	0.938	0.901
	(-0.21)	(1.54)	(2.21)	(-0.87)	(-1.96)				

Table 11 summarizes and compares the conditional alphas of portfolios formed on the basis of alternative cut-offs. Observing the results of these tables, we conclude that forming portfolios that are more strict towards ESG ratings does not change the main results obtained previously: there is no statistical difference between the financial performance of the best and worst rated portfolios whatever the cut-off considered. These results are in constrast to Kempf and Osthoff (2007), who find a positive alpha of the long-short portfolio when using a 10% cut-off but they are in line with those of Halbritter and Dorfleitner (2015), who find no statistically significant alphas when considering alternative cut-offs. A possible explanation for this difference is found below through the errors-in-expectations hypothesis and the shunned-stock hypothesis of Derwall *et al.* (2011).

Table 11. Comparison of conditional alphas between alternative cut-offs (positive screening strategy)

This table summarizes the comparison of conditional alphas (expressed in percentage) between alternative cut-offs. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis.

	Env. Score	Soc. Score	Gov. Score	ESG Score
50% cut-off				
High-rated	0.054*	0.074**	0.047	0.059**
	(1.97)	(2.47)	(1.57)	(2.08)
Low-rated	0.027	0.027	0.032	0.018
	(0.89)	(0.87)	(1.42)	(0.58)
Long-short	0.027	0.047	0.015	0.040
	(0.87)	(1.36)	(0.57)	(1.23)
25% cut-off				
High-rated	0.030	0.077*	0.009	0.058*
	(1.12)	(1.82)	(0.27)	(1.70)
Low-rated	0.013	0.037	0.056*	0.049
	(0.25)	(0.76)	(1.73)	(0.77)
Long-short	0.017	0.041	-0.047	0.009
	(0.33)	(0.58)	(-1.25)	0.14
10% cut-off				
High-rated	0.041	0.088*	-0.024	-0.009
	(1.35)	(1.69)	(-0.52)	(-0.24)
Low-rated	0.011	0.065*	0.051	0.017
	(0.11)	(1.76)	(0.89)	(0.12)
Long-short	0.029	0.024	-0.074	-0.026
	(0.29)	(0.42)	(-1.12)	(-0.21)

5.1.2.2 Correction for outliers

According to Brooks (2014), outliers might affect coefficient estimates because of an increased RSS and, consequently, a decreased R^2. Such observations appear in the tails of the distribution and, consequently, the values of kurtosis are very large. The results presented in appendix F are consistent with this situation. Although it seems that this effect is diluted when forming portfolios with a 50% cut-off, the same does not occur with a 10% cut-off since the number of bonds is lower.

In order to verify if the results are driven by outliers, we also run the conditional model after removing them. Following Edmans (2011), we use the winsorization approach where the lower- and upper-tail are each replaced by the value of the nearest observation to be retained unchanged (Barnett and Lewis, 1984). According to Barnett and Lewis (1984), it is difficult to choose the approriate extent in this method. Brooks (2014) mentions that the removal of outliers is likely to be used to increase artificially the R^2 and it is important to know that each obseration represents useful information. In order to minimize the impact of this approach, the sample is winsorized at the 1th and 99th percentiles.

Appendix I provides the regression results of the multi-factor model for portfolios formed on the basis of individual and combined ESG scores (high- and low-rated portfolios as well as the long-short portfolio) with a 50%, 25% and a 10% cut-off without the outliers. Table 12 summarizes and compares the conditional alphas obtained with alternative cut-offs and excluding outliers. Observing the results of these tables, although the alphas of the low-rated portfolios become statistically significant, we conclude that forming portfolios after excluding outliers does not change the main results obtained previously: there is no statistical difference between the financial performance of the best and worst rated portfolios.

Table 12. Comparison of conditional alphas between alternative cut-offs (positive screening strategy) - without outliers

This table summarizes the comparison of conditional alphas (expressed in percentage) between alternative cut-offs without outliers. ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis.

	Env. Score	Soc. Score	Gov. Score	ESG Score
50% cut-off				
High-rated	0.060**	0.062*	0.061**	0.062**
	(2.03)	(1.95)	(2.11)	(2.17)
Low-rated	0.049**	0.050**	0.041*	0.045**
	(2.37)	(2.13)	(1.78)	(2.26)
Long-short	0.011	0.012	0.021	0.018
	(0.42)	(0.38)	(0.73)	(0.81)
25% cut-off				
High-rated	0.030	0.080*	0.024	0.070**
	(1.13)	(1.90)	(0.74)	(2.04)
Low-rated	0.052**	0.069**	0.063**	0.098***
	(2.43)	(2.31)	(2.19)	(3.13)
Long-short	-0.023	0.011	-0.039	-0.028
	(-0.98)	(0.24)	(-1.17)	(-0.97)
10% cut-off				
High-rated	0.041	0.094*	0.007	0.011
	(1.42)	(1.82)	(0.16)	(0.28)
Low-rated	0.078*	0.059	0.083**	0.100**
	(1.84)	(1.40)	(2.24)	(2.49)
Long-short	-0.037	0.035	-0.077*	-0.090*
	(-0.86)	-0.68	(-1.66)	(-1.97)

5.1.2.3 Performance of equally-weighted portfolios

Some managers prefer an equally-weighted perspective, while others prefer a value-weighted perspective in the market (Christopherson *et al.*, 2009). To take into account these

differences in perspectives, we additionally compute the performance of equally-weighted portfolios.[25]

Table 13 presents the regression results of the conditional multi-factor model for high- and low-rated portfolios as well as the long-short portfolios considering equally-weighted portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off. In general, the results show a higher performance of high-rated and low-rated portfolios. This result suggests that the higher financial performance of the low-rated portfolios is more concentrated in small companies. Nevertheless, the results continue to show that the alphas of the long-short portfolios are not statistically significant. This result is in line with the findings of Halbritter and Dorfleitner (2015), who also find statistically insignificant differences that are robust for different portfolio weightings procedures. It is also worth mentioning that our results are consistent with those of Kempf and Osthoff (2007), who find similar performance results for value- and equally-weighted portfolios. However, Statman and Glushkov (2009) find some differences in the statistical significance of the abnormal returns of long-short portfolios of equally-weighted and value-weighted portfolios.

In conclusion, the main results obtained previously suggesting that there is no statistical difference between the financial performance of the best and worst rated portfolios are robust to the different robustness tests performed.

---

[25] The difference in the weighting scheme between a portfolio and an index directly affects returns (Christopherson *et al.*, 2009). Following several studies that evaluate equally-weighted portfolios using benchmark indices that are value weighted (e. g., Statman and Glushkov, 2009) despite forming equally-weighted portfolios, we continue with the same market index as before.

Table 13. The performance of portfolios formed on individual and combined ESG scores (positive screening strategy) - equally-weighted portfolios

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). *Bond* corresponds to the monthly excess returns of the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are equally-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, **, * indicate significance at the 1%, 5%, and 10% level, respectively. Alphas are expressed in percentage. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.075***	0.807***	0.027**	-0.007	0.001	0.922	0.074	0.071	0.055
	(3.26)	(31.79)	(2.17)	(-0.20)	(0.16)				
Low-rated	0.071**	0.791***	-0.037**	0.047	0.023**	0.817	0.963	0.495	0.634
	(2.13)	(23.03)	(-2.03)	(0.99)	(2.36)				
Long-short	0.004	0.016	0.064***	-0.054	-0.022**	0.184	0.420	0.271	0.367
	(0.12)	(0.46)	(3.42)	(-1.10)	(-2.19)				
Soc. Score									
High-rated	0.074***	0.822***	0.014	0.000	0.002	0.921	0.024	0.015	0.025
	(3.01)	(32.66)	(1.09)	(-0.01)	(0.28)				
Low-rated	0.074**	0.768***	-0.021	0.043	0.019**	0.827	0.748	0.508	0.460
	(2.29)	(23.20)	(-1.20)	(0.95)	(2.07)				
Long-short	0.001	0.053	0.035*	-0.044	-0.018*	0.177	0.032	0.003	0.004
	(0.01)	(1.49)	(1.85)	(-0.89)	(-1.73)				
Gov. Score									
High-rated	0.071***	0.819***	0.009	0.003	0.016	0.854	0.064	0.319	0.000
	(3.04)	(26.98)	(0.55)	(0.08)	(1.42)				
Low-rated	0.068***	0.762***	-0.018	0.032	0.007	0.895	0.217	0.108	0.088
	(3.06)	(32.56)	(-1.24)	(0.97)	(0.87)				
Long-short	0.003	0.057*	0.027	-0.029	0.009	0.170	0.090	0.002	0.000
	(0.10)	(1.79)	(1.12)	(-0.69)	(0.52)				
ESG Score									
High-rated	0.081***	0.835***	0.019	0.021	0.002	0.917	0.083	0.171	0.072
	(3.09)	(31.75)	(1.29)	(0.56)	(0.37)				
Low-rated	0.080**	0.744***	-0.037**	0.034	0.023**	0.824	0.491	0.280	0.300
	(2.58)	(23.26)	(-2.18)	(0.76)	(2.56)				
Long-short	0.001	0.091***	0.056***	-0.013	-0.021**	0.222	0.111	0.216	0.088
	(0.04)	(2.76)	(3.20)	(-0.27)	(-2.21)				

5.1.3 EVOLUTION OF SOCIAL AND FINANCIAL PERFORMANCE OVER TIME

This section addresses the evolution of social and financial performance over time. We start by analyzing how the social ratings of the portfolios evolve over time. Next, motivated by empirical evidence that seems to suggest time-dependency of SRI portfolio performance, we analyze the performance results obtained from the analysis of different sub-periods and market states.

5.1.3.1 Consistency of social ratings over time

Figure 1 shows the evolution of the mean ESG ratings of the socially screened portfolios between 2003 and 2016. We observe that the portfolios formed by bonds with high ethical scores show some consistency in their ESG levels over time. It is important to keep in mind that the rebalancing of these portfolios reflects solely the social ratings of their underlying securities. This constrasts with the rebalancing strategy of most actively managed SRI mutual funds, that rely on fund managers' skills to shift the portfolios' composition in response to a changing market conditions (Auer, 2016; Auer and Schuhmacher, 2016). Hence, there might be a trade-off between keeping the social level of the fund high or taking advantage of market timing and selectivity opportunities. In fact, as Wimmer (2013) points out, the lack of long-term ESG persistence in actively managed mutual funds can be attributed to two reasons. Firstly, an SRI mutual fund manager can change the portfolio's composition according to his investment strategy. Secondly, each companies's ESG scores can change depending on its actions with respect to environmental, social, and governance issues. In his empirical study, Wimmer (2013) concludes that the lack of long-term persistence in mutual funds ESG scores is driven mainly by changes in the holdings of the SRI mutual funds and not by changes in the underlying companies' ESG ratings. The author constructs four equally weighted quartile portfolios of funds (portfolios 1 and 4 contain the 25 percent of all funds with the highest and lowest social ratings in a specific year, respectively) and concludes that there is a significant persistence of the ESG score ranking for the first two years subsequent to their formation. However, there is no empirical evidence of persisting social ratings after

three and four years. This conclusion is relevant for value-driven investors because they want to incorporate CSR practices into their portfolio selection process (Derwall *et al.,* 2011).

Figure 1 also provides additional insights. The low-rated portfolios formed on the basis of the environmental and social scores appear to present a downward trend after the financial crisis of 2007-2008. Even so, after 2014 these ethical ratings of these companies appear to show a recovery. The low-rated portfolios formed on the corporate governance score tend to improve their scores over time. This is consistent with the argument of Lucey and Zhang (2010) that the increasing financial integration over time has allowed firms to borrow funds in countries with more efficient legal systems. In particular, low-rated companies have a financial incentive to improve on corporate governance issues.

Figure 1. Portfolio ESG ratings over time

For each month between 2003 and 2016, the upper (lower) half of each graph in this figure shows the mean ESG scores of portfolios containing the high-rated (low-rated) firms. Portfolios with 50%, 25% and 10% cut-off rates are represented by solid, dashed and dashed-dotted lines, respectively.
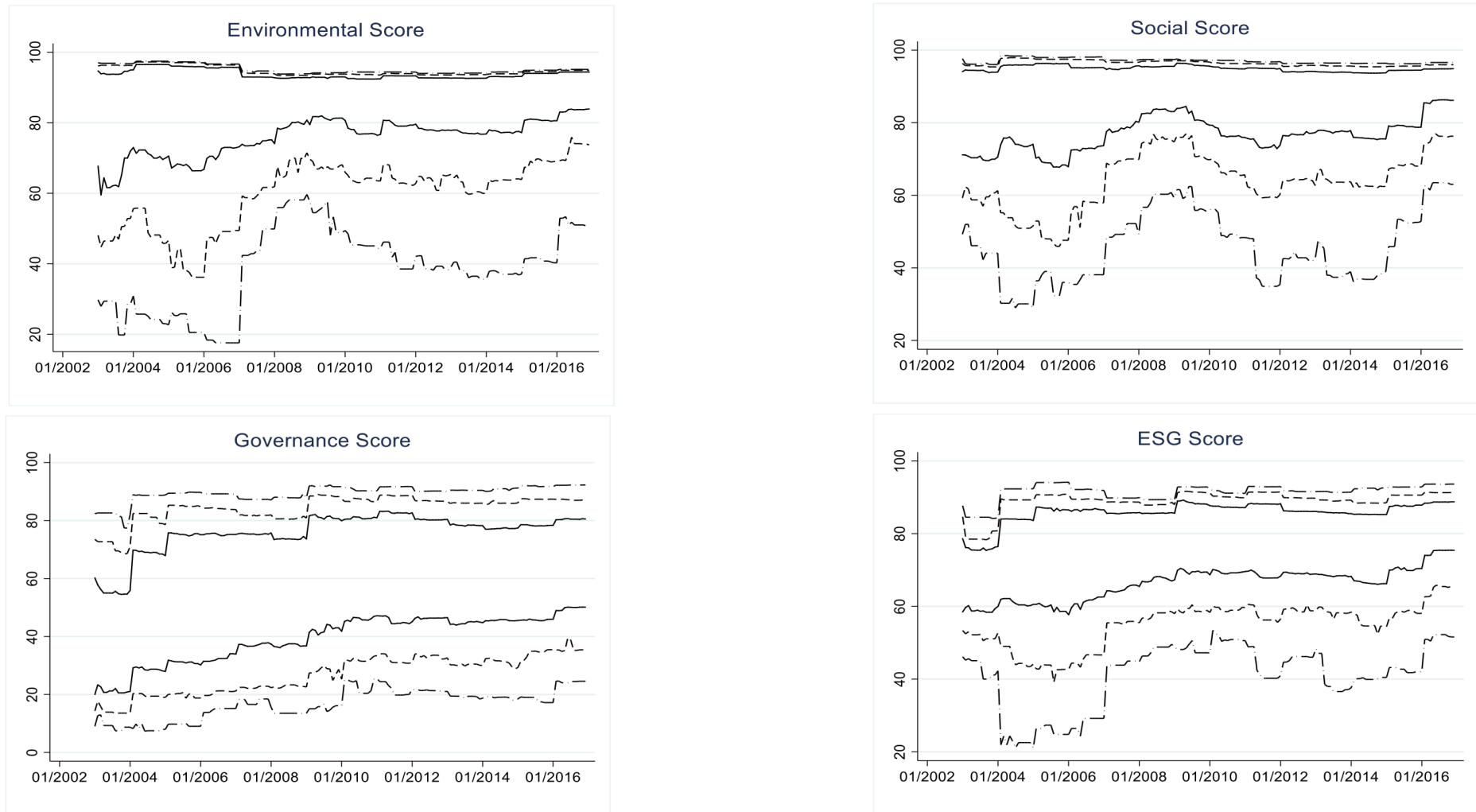
5.1.3.2 SRI bond financial performance over time

According to previous studies, there is a link between ESG scores and financial returns in earlier years (Halbritter and Dorfleitner, 2015). However, investors might no longer expect abnormal returns by trading a portfolio formed on the basis of ESG scores because CSR practices have become recognized as value-relevant by investors in the more recent years (e.g., Derwall *et al.*, 2011; Edmans, 2011; Halbritter and Dorfleitner, 2015). To investigate the performance of the portfolios over time, we follow an approach that is similar to the one used by Derwall *et al.* (2011), by expanding the regression window by 1 year starting with the period 2003-2007[26] and finishing with the period 2003-2016. We choose to apply the full conditional specification of the multi-factor model (equation 10) for the reasons presented before.

Table 14 summarizes the regression results of the conditional multi-factor model for high-rated, low-rated and long-short portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off, over the sub-periods between 2003-2007 and 2003-2016. Except for portfolios formed on the corporate governance score, all long-short portfolios present statistically significant alphas (at the 1% level) during the first period (2003 to 2007), indicating an outperformance of portfolios that score high on ESG ratings compared to those that are less socially responsible. These results indicate that it was possible for investors to obtain abnormal returns by going long in fixed income securities of high-rated companies and short in low-rated companies during this period. Yet, over time the positive alphas of the long-short diminish and lose statistical significance. From 2003 to 2010 the long-short alphas are still positive and statistically significant at the 5% level for portfolios based on environmental and social scores and only at the 10% level for the portfolio based on the ESG score. After this period, high- and low-rated portfolios start to show a performance that is not statistically different. In addition, the positive alphas of high-rated portfolios seem to diminish over time, especially after the period 2003-2010. These results are in line with the findings of Derwall *et al.* (2011) and Edmans (2011). In terms of the corporate governance score, the fact that we do not find evidence of statistically significant abnormal returns on

---

[26] The first period finishes in 2007 because this sub-sample period comprises the minimum number of observations for not observing multicollinearity between the SRI variables.

portfolios formed on this dimension even for the first period (2003-2007) is not surprising. Indeed, it is worth pointing out that Bebchuk *et al.* (2013) find evidence of abnormal returns on portfolios stocks of well-governed companies only prior to 1999. After 2000, there is no evidence of abnormal returns associated to good corporate governance practices. These results are consistent with Gompers *et al.* (2003), who also find statistically significant abnormal performance of portfolios of good governance firms during the 1990s. According to the errors-in-expectations hypothesis of Derwall *et al.* (2011)*,* socially responsible stocks may have higher risk-adjusted returns when the market does not immediately incorporate the value of CSR on expected cash flows. However, the authors expect any abnormal returns associated to errors-in-expectations to be temporary, since in the long run investors should recognize CSR information as a source of companies' intrinsic value. After investors becoming aware that portfolios of well-governed firms yield abnormal returns, these abnormal returns tend to disappear. The results of this study seem to confirm that the errors-in-expectations hypothesis is not only useful to explain the performance of equity portfolios but it is also useful to explain the performance of fixed-income securities over time. In addition, the results of the high-rated portfolios seem to suggest that the price of fixed-income securities of high-rated companies tends to increase over time following increased demand by values-driven investors.

It is also worth pointing out that the alphas of low-rated portfolios appear to increase over time. The alphas of the low-rated portfolios are negative and statistically significant during the period 2003-2007 (except for the corporate governance score), but they turn out to be not statistically different from zero over time. In addition to the errors-in-expectations hypothesis presented before, Derwall *et al.* (2011) also develop the shunned-stock hypothesis that claims that values-driven investors shun socially controversial stocks and hence these stocks will generate higher returns. The results of the low-rated portfolios seem to suggest that values-driven investors are increaslingly shunning fixed-income securities of low-rated companies and, hence, these portfolios generate higher returns over time.

Table 14. The performance of long-short portfolios formed on individual and combined ESG scores with a 50% cut-off (positive screening strategy) - analysis for expanding windows

This table presents estimates of monthly abnormal returns (alphas expressed in percentage) obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy).The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2002 to 2016 based on sub-samples (2003-2007, 2003-2008, 2003-2009, 2003-2010, 2003-2011, 2003-2012, 2003-2013, 2003-2014, 2003-2015 and 2003-2016). ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
Env. Score										
High-rated	0.097	0.056*	0.103***	0.101***	0.076*	0.069**	0.069**	0.061**	0.063**	0.054*
	(1.62)	(1.69)	(2.70)	(2.72)	(1.96)	(2.04)	(2.28)	(2.04)	(2.26)	(1.97)
Low-rated	-0.139**	-0.047	0.009	0.004	0.016	0.021	0.024	0.027	0.031	0.027
	(-2.26)	(-0.80)	(0.15)	(0.08)	(0.35)	(0.51)	(0.63)	(0.74)	(0.93)	(0.89)
Long-short	0.235***	0.103*	0.094*	0.097**	0.060	0.048	0.045	0.034	0.032	0.027
	(3.73)	(1.77)	(1.87)	(2.04)	(1.10)	(0.95)	(1.00)	(0.78)	(0.78)	(0.87)
Soc. Score										
High-rated	0.114***	0.086	0.127***	0.135***	0.102***	0.091***	0.089***	0.084**	0.086**	0.074**
	(3.24)	(1.54)	(2.86)	(3.19)	(2.68)	(2.65)	(2.90)	(2.28)	(2.57)	(2.47)
Low-rated	-0.143**	-0.021	0.025	0.009	0.024	0.030	0.033	0.033	0.031	0.027
	(-2.25)	(-0.34)	(0.42)	(0.17)	(0.50)	(0.69)	(0.86)	(0.92)	(0.92)	(0.87)
Long-short	0.257***	0.108	0.103	0.126**	0.078	0.061	0.057	0.051	0.055	0.047
	(3.56)	(1.53)	(1.64)	(2.09)	(1.50)	(1.30)	(1.18)	(0.98)	(1.30)	(1.36)

Table 14. Continued

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
Gov. Score										
High-rated	0.032	0.030	0.079	0.077	0.064	0.060	0.059	0.051	0.055*	0.047
	(0.60)	(0.63)	(1.51)	(1.61)	(1.49)	(1.53)	(1.59)	(1.45)	(1.68)	(1.57)
Low-rated	-0.032	-0.014	0.033	0.029	0.028	0.028	0.033	0.037	0.038	0.032
	(-0.94)	(-0.43)	(0.99)	(0.88)	(0.95)	(1.02)	(1.33)	(1.43)	(1.63)	(1.42)
Long-short	0.064	0.044	0.045	0.048	0.036	0.031	0.026	0.014	0.017	0.015
	(0.89)	(0.81)	(0.66)	(0.78)	(0.80)	(0.77)	(0.72)	(0.44)	(0.58)	(0.57)
ESG Score										
High-rated	0.090***	0.046	0.100**	0.099**	0.075**	0.067**	0.066**	0.059*	0.067**	0.059**
	(2.85)	(1.32)	(2.48)	(2.60)	(2.14)	(2.09)	(2.01)	(1.89)	(2.25)	(2.08)
Low-rated	-0.157**	-0.035	0.003	-0.003	0.012	0.020	0.023	0.027	0.023	0.018
	(-2.23)	(-0.53)	(0.06)	-0.05	(0.25)	(0.44)	(0.59)	(0.73)	(0.66)	(0.58)
Long-short	0.247***	0.081	0.097	0.102*	0.062	0.047	0.043	0.032	0.044	0.040
	(3.48)	(1.19)	(1.66)	(1.88)	(1.30)	(1.07)	(1.09)	(0.85)	(1.18)	(1.23)

To further explore the differences between the period in which high-rated portfolios outperform and the period where outperformance disappears, we also compare portfolios' performance in two mutually exclusive subperiods. Table 15 provides the regression results of the conditional multi-factor model for portfolios formed on the basis of individual and combined ESG scores (high- and low-rated portfolios as well as the long-short portfolio) with a 50% cut-off, over the periods of 2003 to 2007 (panel A) and 2008 to 2016 (panel B).

Again, we emphazise that all long-short portfolios present statistically significant alphas (at the 1% level) during the first period (2003 to 2007). Yet, over the 2008-2016 period, the performance of both portfolios is not statistically different. This suggests that in an earlier stage the performance of portfolios with high social ratings was higher than the performance of lower rated portfolios, but this overperformance disappeared in a more recent period. While the positive alphas of high-rated portfolios appear to diminish over time, the alphas of low-rated portfolios appear to increase between the two periods (except for the corporate governance score).

The adjusted R^2 for the high- and low-rated portfolios ranges between 0.87 and 0.97 over the 2003-2007 period and between 0.90 and 0.93 over the 2008-2016 period. The results of the Wald test allow us to reject the hypothesis of the conditional alphas being equal to zero (at least at the 5% level) for high- and low-rated portfolios over the 2003-2007 period (except for the high-rated portfolio formed on the corporate governance) and for high- and low-rated portfolios over the 2008-2016 period (except for the high- and low-rated portfolios formed on the basis of the environmental and social scores, respectively). Regarding conditional betas, the results of the Wald test suggest rejecting the null hypothesis that these coefficients are equal to zero for all high- and low rated portfolios over the 2003-2007 period. The results of the Wald test also suggest rejecting the null hypothesis that conditional betas are equal to zero for all low-rated portfolios over the second period and for the high-rated portfolio formed on the basis of the social score. Finally, the Wald test suggests rejecting the hypothesis of the conditional alphas and betas being jointly equal to zero for all high- and low-rated portfolios over the 2003-2007 period. The results of the Wald test also suggest rejecting the hypothesis of the conditional alphas and betas being jointly equal to zero for all low-rated portfolios as well as the high-rated portfolio formed on the basis of the governance score over the 2008-2016 period.

As expected, bond market risk has a significant impact (at the 1% level) on the excess returns of the high- and low-rated portfolios in both periods. Even so, the results show some differences in investment styles between the two periods. It seems that during the 2003-2007 period, high-rated portfolios formed on the basis of social and ESG scores are significantly more exposed to the bond market factor than low-rated portfolios. In the more recent period, high- and low-rated portfolios present similar exposures to the bond market factor. The exception is the high-rated portfolio formed on the environmental score, that is significantly less exposed to this factor. In addition, although high- and low-rated portfolios present similar exposures to the default factor over the 2003-2007 period, high-rated portfolios are more exposed to the default factor than low-rated portfolios over the second period (except for the corporate governance score). This does sound like over the more recent time period, issuers of speculative grade bonds start to be aware of the positive effect of CSR practices in the cost of debt.

Table 15. The performance of portfolios formed on individual and combined ESG scores with a 50% cut-off (positive screening strategy) - sub-periods

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period of 2003 to 2007 (panel A) and 2008 to 2016 (panel B). ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

PANEL A	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.097	1.019***	0.044	0.112	0.014**	0.972	0.000	0.000	0.000
	(1.62)	(19.38)	(1.69)	(0.66)	(2.15)				
Low-rated	-0.139**	0.846***	0.080**	-2.093***	0.020	0.893	0.018	0.001	0.000
	(-2.26)	(10.11)	(2.59)	(-6.08)	(1.41)				
Long-short	0.0235***	0.173*	-0.036	2.205***	-0.006	0.542	0.014	0.001	0.000
	(3.73)	(2.02)	(-1.13)	(6.24)	(-0.42)				
Soc. Score									
High-rated	0.114***	1.035***	0.032*	0.256	0.011	0.966	0.004	0.010	0.000
	(3.24)	(21.52)	(1.82)	(1.30)	(1.40)				
Low-rated	-0.143**	0.836***	0.095***	-2.099***	0.025*	0.889	0.016	0.001	0.000
	(-2.25)	(9.69)	(2.95)	(-5.92)	(1.72)				
Long-short	0.257***	0.199**	-0.062*	2.355***	-0.013	0.505	0.012	0.006	0.000
	(3.56)	(2.02)	(-1.70)	(5.81)	(-0.82)				
Gov. Score									
High-rated	0.032	0.912***	0.076***	-0.775**	0.031**	0.924	0.128	0.012	0.003
	(0.60)	(12.61)	(2.81)	(-2.61)	(2.58)				
Low-rated	-0.032	1.022***	0.012	-0.418**	-0.015*	0.964	0.009	0.000	0.000
	(-0.94)	(22.04)	(0.67)	(-2.19)	(-1.95)				
Long-short	0.064	-0.110	0.064*	-0.357	0.046***	0.339	0.835	0.047	0.061
	(0.89)	(-1.12)	(1.75)	(-0.88)	(2.81)				
ESG Score									
High-rated	0.090***	1.025***	0.051***	0.053	0.011	0.972	0.007	0.006	0.000
	(2.85)	(23.77)	(3.16)	(0.30)	(1.57)				
Low-rated	-0.157**	0.780***	0.081**	-2.390***	0.025	0.868	0.020	0.000	0.000
	(-2.23)	(8.13)	(2.27)	(-6.06)	(1.56)				
Long-short	0.247***	0.245**	-0.030	2.443***	-0.014	0.523	0.023	0.002	0.000
	(3.48)	(2.52)	(-0.84)	(6.12)	(-0.84)				

Table 15. Continued

PANEL B	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.033	0.796***	0.015	0.020	0.002	0.915	0.082	0.385	0.170
	(1.04)	(24.59)	(0.82)	(0.50)	(0.24)				
Low-rated	0.033	0.905***	-0.044***	0.094***	0.016**	0.930	0.034	0.002	0.006
	(1.21)	(32.72)	(-3.15)	(2.75)	(2.17)				
Long-short	0.000	-0.108***	0.058***	-0.074**	-0.014	0.428	0.248	0.011	0.004
	(0.02)	(-3.65)	(3.58)	(-2.05)	(-1.54)				
Soc. Score									
High-rated	0.049	0.822***	0.001	0.024	0.004	0.900	0.007	0.109	0.082
	(1.41)	(23.59)	(0.03)	(0.55)	(0.40)				
Low-rated	0.042*	0.828***	0.004	0.065*	0.005	0.924	0.159	0.002	0.000
	(1.82)	(24.09)	(0.20)	(1.85)	(0.61)				
Long-short	0.006	-0.006	-0.003	-0.041	-0.001	0.197	0.189	0.016	0.016
	(0.18)	(-0.17)	(-0.16)	(-0.94)	(-0.11)				
Gov. Score									
High-rated	0.032	0.840***	0.021	-0.001	0.001	0.917	0.019	0.046	0.027
	(0.97)	(25.18)	(1.17)	(-0.02)	(0.12)				
Low-rated	0.031	0.847***	-0.053***	0.119***	0.018**	0.928	0.013	0.007	0.007
	(1.13)	(30.47)	(-3.46)	(3.50)	(2.14)				
Long-short	0.001	-0.008	0.074***	-0.120***	-0.017*	0.444	0.014	0.000	0.000
	(0.03)	(-0.25)	(4.41)	(-3.20)	(-1.82)				
ESG Score									
High-rated	0.037	0.820***	0.016	0.022	0.002	0.913	0.026	0.222	0.098
	(1.11)	(24.57)	(0.89)	(0.54)	(0.20)				
Low-rated	0.028	0.872***	-0.055***	0.101***	0.019**	0.933	0.022	0.004	0.009
	(1.04)	(31.90)	(-3.64)	(3.03)	(2.36)				
Long-short	0.009	-0.052*	0.071***	-0.079**	-0.017*	0.399	0.061	0.005	0.001
	(0.28)	(-1.69)	(4.22)	(-2.12)	(-1.89)				

As mentioned before, we observe no statistical difference between the financial performance of the best and worst rated portfolios whatever the cut-off considered. These results are in constrast to Kempf and Osthoff (2007), who find a positive alpha of the long-short portolio when using a 10% cut-off. Table 16 summarizes our regression results using the conditional multi-factor model for long-short portfolios formed on the basis of individual and combined ESG scores with a 25% and 10% cut-off portfolios, over alternative sub-periods between 2003-2007 and 2003-2016. Except for portfolio formed on the corporate governance score, long-short portfolios with a 25% cut-off present higher and statistically significant alphas than portfolios with a 50% cut-off during the first period (2003 to 2007). This result is in line with that of Kempf and Osthoff (2007). Yet, over time the positive alphas of the long-short diminish and lose statistical significance. Although the alphas of long-short porftolios formed on the basis of the environmental and ESG scores with a 10% scores are higher than long-short portfolios with a 50% and 25% cut-offs for the first period (2003 to 2007), the results are not statistically significant. Although the number of observations in the time series of the portfolios is equal, the consequences of a small number of bonds included in some periods are reflected in inflated OLS standard errors for the regression coefficients. In this case, *t*-tests have little power and ability to reject the null hypothesis.[27] Even so, it sounds like the alphas of long-short portfolios with more strict cut-offs are higher during the 2003-2007, but they turn out to be similar or even lower than portfolios with a 50% cut-off over time. This suggests that the errors-in-expectations hypothesis and the shunned-stock hypothesis may have the greatest impact at the extremes of the portfolios over time. To test this hypothesis, we divide the portfolio based on the combined ESG score into 4 groups over time.

---

[27] Type II error is the error of not rejecting a false null hypothesis.

Table 16. The performance of long-short portfolios formed on individual and combined ESG scores with alternative cut-offs (positive screening strategy) - analysis for expanding windows

This table presents estimates of monthly abnormal returns (alphas expressed in percentage) obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2002 to 2016 based on sub-samples (2003-2007, 2003-2008, 2003-2009, 2003-2010, 2003-2011, 2003-2012, 2003-2013, 2003-2014, 2003-2015 and 2003-2016). ***, ** and *indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
Env. Score										
25%	0.264*	0.041	0.044	0.046	0.030	0.002	0.018	0.013	0.021	0.017
	(2.00)	(0.36)	(0.43)	(0.48)	(0.37)	(0.02)	(0.28)	(0.21)	(0.36)	(0.33)
10%	0.494*	0.095	0.096	0.109	0.074	0.021	0.030	0.037	0.033	0.029
	(1.82)	(0.43)	(0.46)	(0.57)	(0.45)	(0.14)	(0.25)	(0.31)	(0.30)	(0.29)
Soc. Score										
25%	0.364***	0.117	0.165	0.181*	0.135	0.078	0.062	0.049	0.064	0.041
	(2.93)	(1.06)	(1.66)	(1.94)	(1.25)	(0.81)	(0.72)	(0.57)	(0.81)	(0.58)
10%	0.170*	0.107	0.124	0.135	0.102	0.065	0.037	0.041	0.050	0.024
	(1.69)	(1.35)	(1.41)	(1.70)	(1.45)	(0.99)	(0.63)	(0.61)	(0.82)	(0.42)
Gov. Score										
25%	-0.049	-0.038	-0.038	-0.048	-0.046	-0.044	-0.053	-0.066	-0.044	-0.047
	(-0.75)	(-0.68)	(-0.52)	(-0.80)	(-0.86)	(-1.08)	(-1.41)	(-1.56)	(-1.00)	(-1.25)
10%	-0.243*	-0.257**	-0.191*	-0.158*	-0.174**	-0.178**	-0.159**	-0.148**	-0.079	-0.074
	(-1.75)	(-2.42)	(-1.93)	(-1.69)	(-2.04)	(-2.27)	(-2.28)	(-2.28)	(-1.11)	(-1.12)
ESG Score										
25%	0.372**	0.076	0.087	0.111	0.047	0.018	0.010	0.005	0.019	0.009
	(2.25)	(0.51)	(0.67)	(0.95)	(0.46)	(0.20)	(0.13)	(0.07)	(0.28)	(0.14)
10%	0.554	0.039	0.036	0.098	0.050	0.000	0.007	-0.011	-0.003	-0.026
	(1.47)	(0.13)	(0.13)	(0.40)	(0.24)	(0.00)	(0.04)	(-0.08)	(-0.02)	(-0.21)

Table 17 summarizes the regression results of the conditional multi-factor model for the quartile portfolios formed on the basis of the combined ESG score over the sub-periods between 2003-2007 and 2003-2016. Bonds that have the highest 25% of ESG scores comprise quartile 1 (Q1), whereas bonds that have the lowest 25% of ESG scores comprise quartile 4 (Q4). The top quartile (Q1) presents statistically significant alphas (at the 1% level) during the 2003-2010 period, but afterwards the positive alpha diminishes and loses statistical significance. Surprisingly, the alpha of the second quartile (Q2) increases over time and turns out to be statistically significant. It sounds like the errors-in-expectations hypothesis is only useful to explain the top quartile. Although the bottom quartile presents a lower alpha than the low-rated portfolio with a 50% cut-off, it is not statistiscally significant. As mentioned previously, the consequences of a small number of bonds included in some periods are reflected in inflated OLS standard errors for the regression coefficients. As a consequence, *t*-tests may have little power and ability to reject the null hypothesis. Even so, it is possible to observe that the alpha of the bottom quartile (Q4) is lower than the quartile 3 in the 2003-2007 period but it turns out to be higher over time. This result suggests like the shunned-stock hypothesis is more useful to explain the risk-adjusted returns of the bottom quartile. These results might explain why the performance of portfolios with alternative cut-offs do not match those of Kempf and Osthoff (2007), for example. However, we are aware of the small sample size of the portfolios in some periods as a limitation for the analysis. Hence, further research should consider a wider sample of companies to test the errors-in-expectations hypothesis and the shunned-stock hypothesis with alternative cut-offs and analyze whether both hypotheses have, in fact, the greatest impact at the extremes of the portfolios over time.

Table 17. The performance of portfolio formed on the combined ESG score divided into quartiles (positive screening stragey) - analysis for expanding windows

This table presents estimates of monthly abnormal returns (alphas expressed in percentage) obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The portfolio is value-weighted and rebalanced monthly. The portfolio is divided into quartiles based on the ESG score. Bonds that have the highest quartile of ESG score comprise quartile 1 (Q1), whereas bonds that have the lowest quartile of ESG score comprise quartile 4 (Q4). The observation period spans the period from 2003 to 2016 based on sub-samples (2003-2007, 2003-2008, 2003-2009, 2003-2010, 2003-2011, 2003-2012, 2003-2013, 2003-2014, 2003-2015 and 2003-2016). ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
ESG Score										
Q1	0.123	0.054	0.119**	0.119**	0.087*	0.076*	0.074*	0.063	0.068*	0.058*
	(1.35)	(1.16)	(2.28)	(2.47)	(1.94)	(1.87)	(1.82)	(1.61)	(1.89)	(1.70)
Q2	0.003	0.011	0.047	0.045	0.043	0.042	0.045**	0.045**	0.056***	0.052**
	(0.04)	(0.27)	(1.28)	(1.32)	(1.55)	(1.63)	(2.00)	(2.11)	(2.66)	(2.54)
Q3	-0.064	-0.034	-0.002	0.001	0.004	0.007	0.010	0.018	0.015	0.008
	(-1.58)	(-0.99)	(-0.05)	(0.03)	(0.15)	(0.24)	(0.41)	(0.74)	(0.65)	(0.36)
Q4	-0.249	-0.023	0.032	0.008	0.040	0.057	0.064	0.058	0.049	0.049
	(-1.48)	(-0.15)	(0.24)	(0.07)	(0.38)	(0.63)	(0.83)	(0.78)	(0.71)	(0.77)

5.1.3.3 Portfolio performance in times of crisis

An alternative approach to condition fund performance to the economy involves using dummy variables to distinguish different market states, as in Moskowitz (2000), Kosowski (2006) and Areal *et al.* (2013). Whereas the model of Christopherson *et al.* (1998) conditions performance and risk to the state of the economy by means of continuous public information variables, the dummy variable model presented in equation (13) conditions performance and risk to market states, such as expansion and recession periods. We use a dummy variable which assumes the value of 0 in expansion periods and 1 in recession periods (according to the CEPR) and thus the model enables us to verify if there are statistically differences in the performance and risk of SRI portfolios during "good times" and "bad times".

Table 18 presents the regression results of the dummy variable model of high- and low-rated portfolios as well as long-short portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off.

In comparison with the results obtained with the conditional model of Christopherson *et al.* (1998) specified in a multi-factor context (equation 10), the adjusted R^2 is similar. The results of the Wald test do not allow the rejection of the hypothesis of the incremental alphas in recessions being equal to zero. Regarding incremental betas in recessions, the results of the Wald test suggest rejecting the null hypothesis that these coefficients are equal to zero for the low-rated portfolio formed on the basis of the social score (at the 5%). Finally, the Wald test also suggests rejecting the hypothesis of the incremental alphas and betas being jointly equal to zero for the low-rated portfolio formed on the basis of the social score (at the 5% level).

During expansion periods, the alphas of high- and low-rated portfolios are not statistically different from zero. During recession periods, none of the portfolios changes performance in a statistically significant way. Furthermore, high- and low-rated portfolios perform similar in expansion periods, regardless of the score used, and this does not change in recession periods. As expected, bond market risk has a significant impact (at the 1% level) on the abnormal returns of the high- and low-rated portfolios, regardless of the score used. The results of these portfolios suggest a lower-than-average sensitivity to the market index,

since beta is always lower than 1. It is worth mentioning that with the exception of the social score, all high-rated portfolios are more exposed to the default factor in expansions than low rated portfolios. The exposure of the portfolios formed on the environmental and social scores to this factor is even reinforced in periods of recession, compared to low-rated portfolios. It is possible to observe in appendix E that the percentage of speculative grade bonds is higher for high- than for low-rated portfolios (except for the portfolios based on the social score). As mentioned before, since bonds with low credit ratings present higher yields, issuers of speculative grade bonds can benefit the most in absolute terms from the reductions in the cost of debt that may result from CSR practices (Oikonomou *et al.*, 2014). It does sound like these issuers have a financial incentive to improve on CSR practices.

As mentioned previously in the literature review, there are theoretical arguments in favour a higher performance of SRI firms in times of turmoil. As Hoepner *et al.* (2016) argue, companies that are highly committed to CSR practices can generate reputational wealth and relational capital that prevents declines in value during market crises. Several studies on SRI equity funds (e.g., Nofsinger and Varma, 2014; Leite and Cortez, 2015), on SRI bond funds (e. g., Henke, 2016) and on equity portfolios (e. g., Carvalho and Areal, 2016) find that SRI funds and portfolios provide additional protection during periods of market crisis. Our results do not show evidence that high-rated portfolios provide significant additional protection to investors during recession periods. Even so, it seems that it is possible to invest in portfolios of bonds issued by companies with high ESG ratings without sacrificing the financial performance of investors.

Table 18. The performance of portfolios formed on individual and combined ESG scores (positive screening strategy) - dummy variable model

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the dummy variable model regressions (equation 13). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. D_t is a dummy variable which assumes the value of 0 in expansion periods and 1 in recession periods. The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the incremental alpha in recession, incremental betas in recession and incremental alphas and betas in recession, respectively, are jointly equal to zero.

	α	α_{D_t}	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	β_{Bond*D_t}
Env. Score							
High-rated	0.033	0.040	0.877***	-0.002	0.051	0.004	0.022
	(1.52)	(0.47)	(27.53)	(-0.14)	(1.32)	(0.48)	(0.28)
Low-rated	0.006	0.109	0.909***	-0.029*	0.049	0.016*	-0.034
	(0.21)	(1.57)	(25.83)	(-1.79)	(0.96)	(1.90)	(-0.55)
Long-short	0.027	-0.069	-0.032	0.027**	0.001	-0.012	0.056
	(0.83)	(-1.39)	(-0.75)	(2.02)	(0.04)	(-1.10)	(0.89)
Soc. Score							
High-rated	0.035	0.130	0.890***	-0.022	0.046	0.004	-0.036
	(1.39)	(1.42)	(20.89)	(-1.34)	(1.27)	(0.62)	(-0.49)
Low-rated	0.014	0.099	0.894***	0.008	0.067	0.017**	-0.075
	(0.47)	(1.40)	(25.00)	(0.47)	(1.28)	(1.96)	(-1.20)
Long-short	0.021	0.031	-0.004	-0.030	-0.021	-0.013	0.039
	(0.50)	(0.37)	(-0.08)	(-1.32)	(-0.40)	(-0.91)	(0.57)
Gov. Score							
High-rated	0.011	0.070	0.900***	0.000	0.058	0.015	0.019
	(0.41)	(0.75)	(30.59)	(0.00)	(1.50)	(1.47)	(0.22)
Low-rated	0.040*	0.001	0.875***	-0.035**	0.042	-0.002	-0.042
	(1.95)	(0.96)	(33.17)	(-2.42)	(1.16)	(-0.24)	(-0.68)
Long-short	-0.030	0.004	0.024	0.035**	0.016	0.017	0.061
	(-1.19)	(0.09)	(0.75)	(2.45)	(0.35)	(1.23)	(0.83)
ESG Score							
High-rated	0.028	0.054	0.889***	0.004	0.055	0.006	0.032
	(1.22)	(0.61)	(29.35)	(0.32)	(1.24)	(0.64)	(0.41)
Low-rated	0.014	0.094	0.885***	-0.043**	0.042	0.017*	-0.064
	(0.43)	(1.25)	(23.20)	(-2.47)	(0.75)	(1.81)	(-0.95)
Long-short	0.014	-0.041	0.004	0.047***	0.013	-0.011	0.096
	(0.43)	(-0.53)	(0.11)	(2.68)	(0.22)	(-1.18)	(1.40)

Table 18. Continued

	β_{Def*D_t}	β_{Opt*D_t}	β_{Eq*D_t}	$Adj.R^2$	W_1	W_2	W_3
Env. Score							
High-rated	-0.001	-0.086	0.001	0.901	0.639	0.580	0.462
	(-0.04)	(-1.01)	(0.07)				
Low-rated	-0.054**	0.055	0.012	0.871	0.119	0.065	0.058
	(-2.23)	(0.70)	(0.70)				
Long-short	0.053***	-0.141**	-0.010	0.185	0.167	0.000	0.000
	(2.87)	(-2.48)	(-0.64)				
Soc. Score							
High-rated	0.004	-0.076	0.005	0.887	0.158	0.727	0.483
	(0.13)	(-0.86)	(0.35)				
Low-rated	-0.055**	0.033	0.009	0.862	0.163	0.016	0.022
	(-2.24)	(0.41)	(0.56)				
Long-short	0.060**	-0.109	-0.004	0.076	0.709	0.083	0.133
	(2.19)	(-1.55)	(-0.25)				
Gov. Score							
High-rated	-0.017	-0.117	-0.008	0.889	0.456	0.224	0.152
	(-0.50)	(-1.33)	(-0.38)				
Low-rated	-0.029	0.090	0.030*	0.909	0.339	0.353	0.470
	(-1.05)	(1.26)	(1.96)				
Long-short	0.012	-0.208***	-0.037**	0.224	0.925	0.000	0.001
	(0.51)	(-3.03)	(-2.03)				
ESG Score							
High-rated	-0.019	-0.083	0.000	0.905	0.544	0.365	0.282
	(-0.57)	(-0.97)	(-0.01)				
Low-rated	-0.032	0.072	0.015	0.844	0.213	0.409	0.464
	(-1.23)	(0.84)	(0.84)				
Long-short	0.013	-0.154*	-0.015	0.156	0.597	0.073	0.123
	(0.48)	(-1.79)	(-0.83)				

5.2. SRI BOND PORTFOLIOS FORMED BY USING A BEST-IN-CLASS STRATEGY

5.2.1 PERFORMANCE OF SRI PORTFOLIOS

This chapter presents the results of value-weighted portfolios consisting of European bonds formed on the basis of a best-in-class strategy. By following this approach portfolios are formed with bonds that have the best ESG scores in a particular industry. As mentioned previously, we considered the ten industry classes based on the ICB.

Portfolio performance is evaluated by means of the conditional multi-factor model with time-varying alphas and betas. Table 19 presents the regression results of the best-in-class portfolios formed on the basis of individual and combined ESG scores (high- and low-rated portfolios as well as the long-short portfolio) with a 50% cut-off.

The adjusted R^2 for the different high- and low-rated portfolios ranges between 0.87 and 0.92. The results of the Wald test allow us to reject the hypothesis of the conditional alphas being equal to zero for the high-rated portfolio formed on the social score and for the low-rated portfolio formed on the basis of the governance score. Regarding conditional betas, the results of the Wald test suggest rejecting the null hypothesis that these coefficients are equal to zero for the high-rated portfolio formed on the basis of the ESG score. Finally, the Wald test suggests rejecting the hypothesis of the conditional alphas and betas being jointly equal to zero for the high-rated portfolios formed on the ESG score.

The results show statistically significant differences in investment styles between high- and low rated portfolios. High-rated portfolios formed on the basis of the governance and ESG scores are significantly more exposed to the bond market factor in the case of portfolios formed with the best-in-class approach. High-rated portfolios are significantly more exposed to the default factor for portfolios formed on the basis of environmental and ESG scores. This result is in line with that of portfolios based on the positive approach.

In terms of performance, the results show that high-rated portfolios formed on the basis of the social and ESG scores yield positive and statistically significant abnormal returns (at the 5% level). The low-rated portfolios formed on the basis of the environmental and

governance scores show a positive and statistically significant performance, but the other low-rated portfolios do not show a performance that is statistically different from zero. The alphas of the long-short portfolios do not show statistically significant results. Hence, the results suggest that investors cannot obtain abnormal returns by going long in high-rated stocks and short in low-rated stocks when using the positive or the best-in-class approach. These results are in line with those of Halbritter and Dorfleitner (2015), but contrast with other studies suggesting abnormal returns of an ESG portfolio strategy on the equity market (Derwall *et al.,* 2005; Kempf and Osthoff, 2007; Statman and Glushkov, 2009).

Table 19. The performance of portfolios formed on individual and combined ESG scores (best-in-class strategy) - conditional multifactor model

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings (alphas expressed in percentage), and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD) The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (best-in-class strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.017	0.860***	0.015	0.026	0.006	0.871	0.347	0.332	0.332
	(0.57)	(27.72)	(0.91)	(0.61)	(0.72)				
Low-rated	0.058**	0.878***	-0.025*	0.080***	0.007	0.903	0.456	0.484	0.634
	(2.22)	(30.02)	(-1.76)	(2.69)	(1.03)				
Long-short	-0.041	-0.018	0.040**	-0.054	-0.001	0.012	0.827	0.993	0.992
	(-1.29)	(-0.55)	(2.33)	(-1.19)	(-0.10)				
Soc. Score									
High-rated	0.064**	0.852***	-0.005	0.055*	0.004	0.894	0.032	0.151	0.154
	(2.14)	(28.68)	(-0.29)	(1.68)	(0.48)				
Low-rated	0.033	0.885***	-0.010	0.060**	0.006	0.915	0.162	0.088	0.123
	(1.27)	(32.01)	(-0.66)	(2.21)	(1.08)				
Long-short	0.031	-0.033	0.005	-0.004	-0.002	0.187	0.004	0.000	0.001
	(1.06)	(-1.33)	(0.25)	(-0.10)	(-0.20)				
Gov. Score									
High-rated	0.033	0.903***	0.007	0.017	0.004	0.914	0.305	0.161	0.234
	(1.34)	(32.26)	(0.43)	(0.50)	(0.57)				
Low-rated	0.062***	0.825***	-0.020	0.082***	0.011	0.911	0.013	0.088	0.072
	(2.66)	(37.10)	(-1.42)	(2.84)	(1.51)				
Long-short	-0.029	0.078***	0.027	-0.064*	-0.007	0.096	0.214	0.593	0.579
	(-1.06)	(3.51)	(1.42)	(-1.68)	(-0.71)				
ESG Score									
High-rated	0.053**	0.877***	0.022*	0.023	-0.001	0.922	0.169	0.030	0.020
	(2.26)	(33.97)	(1.67)	(0.73)	(-0.18)				
Low-rated	0.047*	0.819***	-0.025	0.063*	0.017**	0.889	0.065	0.266	0.302
	(1.77)	(26.58)	(-1.53)	(1.74)	(2.09)				
Long-short	0.006	0.058**	0.047***	-0.040	-0.018**	0.161	0.030	0.035	0.040
	(0.21)	(2.12)	(3.24)	(-1.06)	(-2.34)				

5.2.2 EVOLUTION OF SOCIAL AND FINANCIAL PERFORMANCE OVER TIME

5.2.2.1 SRI bond financial performance over time

Table 20 summarizes the regression results of the conditional multi-factor model for long-short portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off, over the sub-periods between 2003-2007 and 2003-2016. As mentioned previously, we follow an approach that is similar to the one used by Derwall *et al.* (2011).

The long-short portfolios formed on the basis of the social and ESG scores present statistically significant alphas during the first period (2003 to 2007), indicating an outperformance of portfolios that score high on ESG ratings compared to those that are less socially responsible. Yet, over time the positive alphas diminish and high- and low-rated portfolios start to perform similarly. In addition, the positive alphas of high-rated portfolios appear to diminish over time. These results seem to confirm that the errors-in-expectations hypothesis of Derwall *et al.* (2011) is also useful to explain the performance of portfolios formed on the basis of the social and ESG scores when using the best-in-class approach. Although the evidence does not show statistically significant abnormal returns on portfolios formed on the basis of the environmental score when using the best-in-class approach, there is statistically significant results on this dimension when using the positive approach.

The alphas of low-rated portfolios appear to increase over time, except for the portfolio formed on the basis of the corporate governance score. Although it is possible to observe an increasing tendency, the alphas of the low-rated portfolio formed on the social score are never statistically different from zero. In line with results based on the positive approach, these results seem to suggest that values-driven investors shun fixed-income securities of low-rated companies due to the universal nature of certain social values.

Table 20. The performance of long-short portfolios formed on individual and combined ESG scores (best-in-class strategy) - analysis for expanding windows

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), obtained from the conditional multi-factor model regressions (equation 10). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (best-in-class strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2002 to 2016 based on sub-samples (2003-2007, 2003-2008, 2003-2009, 2003-2010, 2003-2011, 2003-2012, 2003-2013, 2003-2014, 2003-2015 and 2003-2016). ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. Alphas are expressed in percentage. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
Env. Score										
High-rated	-0.052	0.021	0.047	0.045	0.039	0.040*	0.039*	0.035*	0.023	0.017
	(-0.93)	(0.48)	(1.13)	(1.07)	(1.07)	(1.71)	(1.85)	(1.87)	(0.74)	(0.57)
Low-rated	0.043	0.029	0.080*	0.074*	0.065*	0.061*	0.064**	0.066**	0.065**	0.058**
	(0.79)	(0.95)	(1.85)	(1.81)	(1.85)	(1.86)	(2.06)	(2.21)	(2.38)	(2.22)
Long-short	-0.095	-0.008	-0.033	-0.029	-0.026	-0.021	-0.024	-0.030	-0.042	-0.041
	(-0.75)	(-0.14)	(-0.66)	(-0.62)	(-0.65)	(-0.58)	(-0.74)	(-0.96)	(-1.25)	(-1.29)
Soc. Score										
High-rated	0.085**	0.046	0.109***	0.111*	0.091**	0.079**	0.079**	0.075**	0.074**	0.064**
	(2.34)	(1.50)	(2.83)	(1.97)	(2.54)	(2.07)	(2.36)	(2.30)	(2.42)	(2.14)
Low-rated	-0.054	-0.018	0.028	0.032	0.031	0.040	0.039	0.039	0.039	0.033
	(-1.19)	(-0.39)	(0.61)	(0.75)	(0.84)	(1.21)	(1.33)	(1.26)	(1.41)	(1.27)
Long-short	0.139**	0.064	0.081	0.079	0.060	0.039	0.040	0.036	0.034	0.031
	(2.31)	(1.36)	(1.16)	(1.07)	(1.05)	(0.78)	(0.87)	(1.05)	(1.08)	(1.06)

Table 20. Continued

	2003-2007	2003-2008	2003-2009	2003-2010	2003-2011	2003-2012	2003-2013	2003-2014	2003-2015	2003-2016
Gov. Score										
High-rated	-0.022	-0.023	0.029	0.032	0.031	0.032	0.036	0.031	0.037	0.033
	(-0.53)	(-0.61)	(0.72)	(0.85)	(0.89)	(1.01)	(1.23)	(1.11)	(1.46)	(1.34)
Low-rated	0.082**	0.079**	0.107**	0.097**	0.084**	0.081**	0.076***	0.078***	0.071***	0.062***
	(2.24)	(2.59)	(2.54)	(2.18)	(2.40)	(2.46)	(2.61)	(2.78)	(2.89)	(2.66)
Long-short	-0.104	-0.102**	-0.078*	-0.065	-0.053	-0.049	-0.040	-0.047	-0.034	-0.029
	(-1.63)	(-2.19)	(-1.73)	(-1.54)	(-1.42)	(-1.43)	(-1.25)	(-1.53)	(-1.14)	(-1.06)
ESG Score										
High-rated	0.104***	0.046	0.094**	0.092**	0.068**	0.062**	0.060**	0.054**	0.063**	0.053**
	(3.23)	(1.47)	(2.59)	(2.62)	(2.12)	(2.15)	(2.35)	(2.13)	(2.57)	(2.26)
Low-rated	-0.038	-0.009	0.029	0.023	0.046	0.045	0.058*	0.065**	0.054*	0.047*
	(-1.15)	(-0.26)	(0.81)	(0.62)	(1.24)	(1.31)	(1.80)	(2.12)	(1.87)	(1.77)
Long-short	0.142***	0.055	0.065	0.069*	0.023	0.016	0.002	-0.010	0.009	0.006
	(3.13)	(1.24)	(1.54)	(1.75)	(0.47)	(0.48)	(0.08)	(-0.34)	(0.28)	(0.21)

5.2.2.2 Portfolio performance in times of crisis

Table 21 presents the regression results of the dummy variable model for high- and low-rated portfolios as well as the long-short portfolios for portfolios formed on the basis of individual and combined ESG scores with a 50% cut-off.

In comparison with the results obtained with the conditional model of Christopherson *et al.* (1998) specified in a multi-factor context (equation 10), the results using a dummy variable show a slight decreasing adjusted R^2. The results of the Wald test do not allow the rejection of the hypothesis of the incremental alphas and betas in recessions being equal to zero.

Regarding differences in investment styles between high- and low-rated portfolios, the high-rated portfolio formed on the basis of the ESG score is more exposed to the default factor in expansions than low-rated portfolios. It is worth mentioning that with the exception of the social score, when portfolios were formed on the positive approach all high-rated portfolios are more exposed to the default factor in expansions than low rated portfolios.

During expansion periods, alphas of high-rated portfolios are not statistically different from zero. Yet, alphas of low-rated portfolios formed on the basis of governance and ESG scores are positive and statistically significant. During recession periods, none of the portfolios changes performance in a statistically significant way. Furthermore, high- and low-rated portfolios perform similar in expansion periods, regardless of the score used, and this does not change in recession periods. These results are in line with those of the portfolios based on the positive approach.

Table 21. The performance of portfolios formed on individual and combined ESG scores (best-in-class strategy) - dummy variable model

This table presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the dummy variable model regressions (equation 13). Bond, Default, Option and Equity represent the risk factors, as described in the previous tables. D_t is a dummy variable which assumes the value of 0 in expansion periods and 1 in recession periods. The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG scores (best-in-class strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively. Alphas are expressed in percentage. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the incremental alpha in recession, incremental betas in recession and incremental alphas and betas in recession, respectively, are jointly equal to zero.

	α	α_{D_t}	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	β_{Bond*D_t}
Env. Score							
High-rated	0.012	0.078	0.834***	-0.002	0.029	0.009	0.054
	(0.39)	(1.13)	(23.77)	(-0.11)	(0.56)	(1.02)	(0.88)
Low-rated	0.045*	0.060	0.892***	-0.030**	0.066	0.006	-0.033
	(1.75)	(1.02)	(29.93)	(-2.18)	(1.50)	(0.84)	(-0.63)
Long-short	-0.033	0.018	-0.058	0.028*	-0.037	0.003	0.088
	(-1.09)	(0.26)	(-1.65)	(1.73)	(-0.71)	(0.30)	(1.40)
Soc. Score							
High-rated	0.038	0.063	0.845***	-0.012	0.025	-0.001	0.014
	(1.57)	(0.67)	(31.41)	(-0.75)	(0.60)	(-0.10)	(0.19)
Low-rated	0.035	0.081	0.897***	-0.013	0.067**	0.011	-0.041
	(1.64)	(0.91)	(38.11)	(-1.00)	(2.16)	(1.51)	(-0.52)
Long-short	0.003	-0.018	-0.052**	0.001	-0.042	-0.012	0.055
	(0.09)	(-0.16)	(-2.27)	(0.05)	(-0.94)	(-0.97)	(0.69)
Gov. Score							
High-rated	0.018	0.071	0.904***	-0.013	0.055	0.008	-0.024
	(0.79)	(0.84)	(35.44)	(-0.94)	(1.34)	(0.94)	(-0.31)
Low-rated	0.055**	0.055	0.824***	-0.016	0.016	0.005	0.010
	(2.56)	(0.96)	(26.56)	(-0.84)	(0.40)	(0.79)	(0.18)
Long-short	-0.037	0.016	0.080***	0.003	0.039	0.003	-0.033
	(-1.44)	(0.27)	(2.66)	(0.22)	(0.89)	(0.35)	(-0.63)
ESG Score							
High-rated	0.037	0.037	0.869***	0.006	0.037	0.001	0.007
	(1.62)	(0.54)	(33.06)	(0.46)	(0.99)	(0.14)	(0.11)
Low-rated	0.052**	0.059	0.812***	-0.030**	0.008	0.009	0.005
	(2.30)	(0.60)	(33.11)	(-2.33)	(0.22)	(1.32)	(0.05)
Long-short	-0.015	-0.022	0.058*	0.036**	0.029	-0.008	0.003
	(-0.56)	(-0.35)	(1.86)	(2.54)	(0.64)	(-1.03)	(0.05)

Table 21. Continued

	β_{Def*D_t}	β_{Opt*D_t}	β_{Eq*D_t}	$Adj.R^2$	W_1	W_2	W_3
Env. Score							
High-rated	-0.015	-0.013	0.005	0.869	0.261	0.796	0.557
	(-0.63)	(-0.17)	(0.29)				
Low-rated	-0.021	-0.005	0.010	0.902	0.312	0.685	0.724
	(-1.01)	(-0.08)	(0.72)				
Long-short	0.005	-0.008	-0.005	0.065	0.793	0.616	0.637
	(0.22)	(-0.10)	(-0.31)				
Soc. Score							
High-rated	-0.046	-0.010	0.019	0.881	0.503	0.387	0.371
	(-1.19)	(-0.09)	(0.89)				
Low-rated	-0.008	0.012	0.002	0.913	0.364	0.904	0.862
	(-0.30)	(0.22)	(0.20)				
Long-short	-0.038	-0.022	0.017	0.074	0.870	0.474	0.595
	(-1.02)	(-0.18)	(0.78)				
Gov. Score							
High-rated	-0.006	-0.111	0.003	0.910	0.402	0.311	0.341
	(-0.19)	(-1.28)	(0.15)				
Low-rated	-0.036	0.095	0.012	0.897	0.338	0.641	0.644
	(-1.48)	(1.35)	(1.03)				
Long-short	0.029	-0.206***	-0.009	0.137	0.788	0.037	0.066
	(1.41)	(-3.07)	(-0.63)				
ESG Score							
High-rated	-0.021	-0.078	0.004	0.915	0.593	0.348	0.447
	(-0.72)	(-1.10)	(0.25)				
Low-rated	-0.023	0.122	0.019	0.877	0.548	0.696	0.812
	(-0.73)	(1.34)	(1.07)				
Long-short	0.002	-0.200***	-0.016	0.148	0.726	0.008	0.017
	(0.09)	(-2.89)	(-1.06)				

5.3. DISCUSSION OF RESULTS

According to Markowitz (1952), the portfolio mean-variance optimization framework implies diversification across companies with different economic and financial characteristics because they have lower convariance between them. Furthermore, the semi-strong form of market efficiency as described by Fama (1970) claims that portfolio managers that use only publicly available information do not generate positive alphas. In contrast, consistent with the stakeholder approach (Freeman, 1984), supporters of the outperformance of socially responsible investing argue that companies that consider the viewpoint of all stakeholders might benefit from long-term higher performance. CSR can be viewed as a source of competitive advantage that could be reflected in higher financial performance and higher shareholder value (Nollet *et al.*, 2016).

There are two sources of value in active bond management. The first is timing the market by changing the duration of a portfolio to benefit from future changes in interest rates (market timing). If managers expect an increase of interest rates, they decrease duration (and vice versa). Several studies argue that bond returns are a function of systematic risk factors, since bonds constitute a homogenous asset class (Derwall and Koedijk, 2009). The second source is the identification of relative mispricing within the fixed-income market (security selection). Portfolio managers might consider that the default premium on a specific bond is too large and, consequently it is underpriced. Indeed, several other studies (e.g., Dynkin *et al.*, 1999; Hottinga *et al.*, 2001; and Dynkin *et al.*, 2002), show that firm-specific risk is also a share of corporate bonds risk and can either be exploited by active management or be eliminated by diversification across several securities and asset classes (Derwall and Koedijk, 2009). This suggests that CSR might also have an impact on bond investment performance.

Indeed, Henke (2016) not only documents positive and statistically significant alphas of SRI bond funds but also that these funds outperform conventional ones (however only at the 10% level in the Eurozone market). According to the errors-in-expectations hypothesis put forward by Derwall *et al.* (2011), financial markets do not incorporate CSR information timely nor efficiently and thus underestimate the importance of ESG issues. In contrast to the proposition of the efficient market hypothesis described by Fama (1970), evidence of positive alphas support the claim that securities prices do not fully reflect all available information

and, hence, bond securities are mispriced. To test whether there is a systematic effect of social screening on financial returns, Henke (2016) conducts multifactor model regressions with an ESG screening factor. After controlling for this factor, the alphas of SRI bond funds decline significantly. In addition, the author performs a performance attribution analysis and concludes that the bond portfolio ESG screening factor explains a portion of bonds' active returns. Hence, the results of Henke (2016) are consistent with the hypotheses of mispriced ESG risks, which can also explain the positive and statistically significant alphas found in high-rated portfolios in this study.

The tendency of increasing alphas of low-rated portfolios over time cannot be explained by the beneficial effects of CSR activities. Yet, they can be explained by the shunned-stock hypothesis of Derwall *et al.* (2011). In addition to values-driven investors who appear to shun bond securities issued by low-rated companies, Henke (2016) argues that several managers of SRI bond funds follow an investment strategy that reduce the fund's exposure to corporate bonds with low social ratings. This worst-in-class exclusion approach also contributes to explain the higher risk-adjusted return of low-rated portfolios over time documented in this study.

Derwall *et al.* (2011) also posit that the errors-in-expectation hypothesis and the shunned-stock hypothesis are explained by a division of the SRI movement into a profit-seeking approach and a values-driven investment approach. While profit-seeking investors seek traditional financial performance, values-driven investors consider social and personal values in their portfolio selection process, being willing to assume a loss in financial performance in exchange for non-financial utility. The authors point out several studies consisting of interviews and surveys (e.g., Beal and Goyen, 1998) and focusing on actual holdings and investment decisions of institutional and retail investors (e.g., Johnson & Greening, 1999) that support the idea that not all investors are alike and that values-driven and profit-seeking socially responsible investors coexist.

Until now we presented arguments that can explain the performance of SRI bond portfolios documented in this study. We now discuss the hypotheses that can explain the differences in results between this study and most of the empirical evidence on actively managed SRI bond funds. Henke (2016) claims that funds incorporating higher average ESG

scores compared to conventional funds can generate a higher performance. However, there is evidence suggesting that the fact that a fund is classified as SRI does not ensure portfolios with ESG ratings above the average of conventional funds (Auer and Schuhmacher, 2016; Henke, 2016). For instance, Wimmer (2013) finds that funds classified as socially responsible considerably change their social standards over time and, hence, they can present lack of long-term ESG persistence. Therefore, the positive effect of social screening on financial returns is not reflected in these funds. This contrasts with the results obtained in this study, since portfolios formed by companies with high ethical scores show consistency in their ESG levels over time. In this dissertation, the portfolios' composition reflect solely the social ratings of their underlying holdings and do not rely on fund managers' investment strategies (Auer, 2016; Auer and Schuhmacher, 2016). SRI mutual fund performance is due to both the skills of the fund manager and to the socially responsible characteristics of the companies held by funds (Kempf and Osthoff, 2007). Hence, there might be a trade-off between keeping the social level of the fund high or trying to taking advantage of market and selectivity opportunities. In addition, SRI bond fund managers can engage in a hybrid strategy of including bond securities issued by high-rated companies and excluding bond securities issued by low-rated companies. The two distinct effects could cancel out, possibly leading to a zero net effect. The "no effect" hypothesis is mentioned by Derwall *et al.* (2011) and by Statman and Glushkov (2009), for example.

In the long run, Derwall *et al.* (2011) do not expect that both hypotheses will coexist. Although errors-in-expectations should disappear as investors improve their understanding of CSR information as a source of companies' intrinsic value, the authors expect that the shunned-stock effect will persist over time due to the nature of certain values, since values-driven investors are willing to assume a loss in financial performance in exchange for non-financial utility. Following this reasoning, we expect that any abnormal returns of high-rated portfolios disappear over time and start performing similar to the market. Conversely, as long as values-driven investors still exist or even increase, we expect that the abnormal returns of low-rated portfolios are maintained or even increase over time.

6. CONCLUSION

This dissertation investigates the performance of socially screened bond portfolios of 189 Eurozone companies between 2003 and 2016. As far as we are aware of, this is the first study to evaluate socially responsible investments by forming bond portfolios on the basis of social ratings of European companies.

We form value-weighted portfolios of bonds based on the companies' social scores provided by ASSET4 ESG database. Portfolios are formed with respect to individual dimensions of CSR (environment, social and governance dimensions) as well as to an aggregate (combined) measure of CSR. The high- and low-rated portfolios consist of the top (above the median score) and bottom (below the median score) of all firms with bonds in each period, respectively. Portfolios are rebalanced monthly to account for any new securities issues or redemptions and changes in social ratings that occur throughout the year. In addition, long-short portfolios (long in the high-rated and short in the low- rated portfolio) are also constructed, to better assess the differences in abnormal returns from investing in high-rated and low-rated portfolios. Portfolio performance is evaluated by using unconditional and conditional model that accounts for four risk factors.

The initial findings of this study suggest that high-rated bond portfolios formed with positive screening strategies present, in general, a performance which is not statistically different from that of low-rated bond portfolios. Our findings are robust to both unconditional and conditional models of performance evaluation, as well as to a different portfolio weighting scheme, alternative cut-off portfolios and the exclusion of outliers. Moreover, this conclusion does not change by constructing portfolios based on the best-in-class strategy.

We also analyze time effects of SRI bond investing. Analyzing the evolution of SRI bond portfolio performance over time showed interesting findings. In fact, the results indicate that in an earlier stage it was possible for investors to obtain abnormal returns by going long in fixed income securities of high-rated companies and short in fixed income securities of low-rated companies. However, over time this outperformance diminishes and loses statistical

significance. Although the positive alphas of high-rated portfolios decrease over time, the alphas of low-rated portfolios exhibit an increasing pattern of returns. The distinct patterns of performance of high- and low-rated bond portfolios are consistent with the errors-in-expectations and the shunned-stock hypotheses of Derwall *et al.* (2011). According to the errors-in-expectations hypothesis, SRI can deliver superior performance if the market does not immediately incorporate the positive impact of CSR on the company's intrinsic value. However, in the long run it is expected that investors recognize the importance of CSR practices and that any evidence of mispricing disappears. In addition, the shunned-stock hypothesis claims that values-driven investors shun socially controversial stocks and hence these stocks will generate higher returns. The results of the low-rated portfolios seem to suggest that there has been an increasing number of values-driven investors that shun fixed-income securities of low-rated companies. In addition, the results of the high-rated portfolios are consistent with both the mispricing argument and the claim that the price of fixed-income securities of high-rated companies tend to increase over time because there has been an increased demand by values-driven investors. The results of this study seem to confirm that the errors-in-expectations hypothesis and the shunned-stock hypothesis are not only useful to explain the performance of equity portfolios but they are also useful in explaining the performance of fixed-income securities over time, especially through portfolios formed on the basis of the ESG, environmental and social scores. However, we do not find evidence that both hypothesis can explain the behavior of portfolios formed on the basis of the corporate governance dimension, neither when using the positive/ESG integration nor best-in-class strategies. This result is not surprising considering the results of Gompers *et al.* (2003) and Bebchuk *et al.* (2013), who find evidence of abnormal returns on portfolios stocks of well-governed companies only prior to 1999.

This paper further explores the performance of SRI portfolios in times of crisis. The results do not show differences in the performance of high- and low-rated portfolios in periods of expansions compared to recessions. Even so, it seems that it is possible to invest in portfolios of bonds issued by companies with high ESG ratings without sacrificing the financial performance of investors.

The results also show statistical significant differences in investment styles between high- and low rated portfolios. High-rated portfolios are significantly more exposed to the

default factor for portfolios formed on the basis of environmental and ESG scores. It is worth mentioning that with the exception of the social score, all high-rated portfolios are more exposed to the default factor in expansions periods. The exposure of the portfolios formed on the environmental and social scores to this factor is even reinforced in periods of recession. Considering that bonds with low credit ratings present high yields, issuers of speculative grade bonds can benefit the most in absolute terms from the reductions in the cost of debt that may arise from CSR practices (Oikonomou *et al.*, 2014). Thus, it does sound like these issuers have a financial incentive to improve on CSR practices.

In conclusion, this paper shows that in an earlier stage it was possible for investors to obtain abnormal returns by going long in fixed income securities of high-rated companies and short in low-rated companies. However, investors should no longer expect abnormal returns. Even so, investors can choose SRI bond portfolios without sacrificing financial performance. Hence, socially responsible investors can do well while doing good.

As in any other study, this study presents some limitations. As the main limitation of this study we point out the small size of the sample. Although this limitation does not seem to affect the results by constructing portfolios with a 50% cut-off, it can affect the results by constructing portfolios with a 10% cut-off, given the small number of bonds included in some periods. Further research should consider a wider sample to test alternative cut-offs as well as the exclusion strategy of some controversial business areas.[28] Further research should also analyze whether the errors-in-expectations hypothesis and the shunned-stock hypothesis have the greatest impact at the extremes of the portfolios over time. This type of analysis can help explain the results of alternative cut-offs found in this dissertation, which contrast with those of Kempf and Osthoff (2007). To the best of our knowledge there is no empirical evidence that considers this type of analysis. As for future research, we suggest considering the impact of transaction costs in the performance of the long-short strategies. It is also important to better understand the influence of ESG screenings on bond portfolios from different geographic regions, credit ratings and bond maturities. Moreover, this study could be applied to the financial industry.

---

[28] For example, weapons, alcohol, tobacco, gambling, etc.

REFERENCES

Adcock, C. J., Cortez, M. C., Armada, M. J., & Silva, F. (2012). Time varying betas and the unconditional distribution of asset returns. *Quantitative Finance, 12*, 951-967.

Areal, N., Cortez, M., & Silva, F. (2013). The conditional performance of US mutual funds over different market regimes: Do different types of ethical screens matter? *Financial Markets and Portfolio Management, 27*(4), 397-429.

Ashbaugh-Skaife, H., Collins, D. W., & LaFond, R. (2006). The effects of corporate governance on firms' credit ratings. *Journal of Accounting and Economics, 42*, 203-243.

Auer, B. R. (2016). Do socially responsible investment policies add or destroy European stock portfolio value? *Journal of Business Ethics, 135*(2), 381-397.

Auer, B., & Schuhmacher, F. (2016). Do socially (ir)responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance, 59*, 51-62.

Ayadi, M. A., Ben-Ameur, H., & Kryzanowski, L. (2016). Typical and tail performance of Canadian equity SRI mutual funds. *Journal of Financial Services Research, 50*(1), 57-94.

Barnett, V., & Lewis, T. (1984). *Outliers in Statistical Data.* Chichester: John Wiley & Sons.

Bauer, R., & Hann, D. (2010). Corporate environmental management and credit risk. [Working Paper] SSRN eLibrary.

Bauer, R., Derwall, J., & Otten, R. (2007). The ethical mutual fund performance debate: New evidence from Canada. *Journal of Business Ethics, 70*(2), 111-124.

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

Bauer, R., Otten, R., & Rad, A. (2006). Ethical investing in Australia: Is there a financial penalty? *Pacific-Basin Finance Journal, 14*, 33-48.

Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata.* Texas: Stata Press.

Beal, D., & Goyen, M. (1998). Putting your money where your mouth is, a profile of ethical investors. *Financial Services Review, 7*, 129-143.

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

Becchetti, L., Ciciretti, R., Dalò, A., & Herzel, S. (2015). Socially responsible and conventional investment funds: Performance comparison and the global financial crisis. *Applied Economics, 47*(25), 2541–2562.

Bello, Z. Y. (2005). Socially responsible investing and portfolio diversification. *The Journal of Financial Research, 28*(1), 41-57.

Bessembinder, H., Kahle, H., Maxwell, W., & Xu, D. (2009). Measuring abnormal bond performance. *The Review of Financial Studies, 22*(10), 4219– 4258.

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments.* New York: McGraw-Hill Education.

Borgers, A., Derwall, J., Koedijk, K., & Horst, J. (2013). Stakeholder relations and stock returns: On errors in investors' expectations and learnings. *Journal of Empirical Finance, 22*, 159-175.

Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: UK evidence from disaggregate measures. *Financial Management*, 97-116.

Breusch, T. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers, 17*(31), 334-355.

Brooks, C. (2014). *Introductory Econometrics for Finance.* Cambridge: Cambridge University Press.

Brzeszczyñski, J., & McIntosh, G. (2014). Performance of portfolios composed of British SRI stocks. *Journal of Business Ethics, 120*, 335–362.

Callan, S., & Thomas, J. (2009). Corporate financial performance and corporate social performance: An update and reinvestigation. *Corporate Social Responsibility and Environmental Management, 17*, 61-78.

Campbell, J. Y., & Taksler, G. B. (2003). Equity volatility and corporate bond yields. *The Journal of Finance, 58*(6), 2321-2350.

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

Carvalho, A., & Areal, N. (2016). Great Places To Work®: Resilience in times of crisis. *Human Resource Management, 55*(3), 479–498.

Christopherson, J. A., Cariño, D. R., & Ferson, W. E. (2009). *Portfolio Performance, Measurement and Benchmarking.* New York: McGraw-Hill.

Christopherson, J., Ferson, W., & Glassman, D. (1998). Conditioning manager alphas on economic information: Another look at the persistence of performance. *The Review of Financial Studies, 11*, 111-142.

Climent, F., & Soriano, P. (2011). Green and good? The investment performance of US environmental mutual funds. *Journal of Business Ethics*, 275-287.

Collison, D., Cobb, G., Power, D., & Stevenson, A. (2008). The financial performance of the FTSE4Good indices. *Corporate Social Responsibility and Environmental Management, 15*, 14-28.

Core, J. E., Guay, W. R., & Rusticus, T. O. (2006). Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations. *The Journal of Finance, 61*(2), 655-687.

Cortez, M. C., Silva, F., & Areal, N. (2009). The performance of European socially responsible funds. *Journal of Business Ethics, 87*(4), 573-588.

Derwall, J., & Koedijk, K. (2009). Socially responsible fixed-income funds. *Journal of Business Finance & Accounting, 36*(1/2), 210-229.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analyst Journal, 61*(2), 51-63.

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

D'Agostino, R. B., Belanger, A. J., & D'Agostino, Jr, R. B. (1990). A suggestion for using powerful and informative tests of normality. *American Statistician, 44*, 316-321.

Drut, B. (2010). Sovereign bonds and socially responsible investment. *Journal of Business Ethics, 92*, 131-145.

Dynkin, L., Ferket, P., Hayman, J., van Leeuwen, E., & Wu, W. (1999). Value of security selection versus asset allocation in credit markets. *Journal of Portfolio Management, 25*, 11-27.

Dynkin, L., Hyman, J., & Konstantinovsky, V. (2002). Sufficient diversification in credit portfolios. *Journal of Portfolio Management, 29*, 89-114.

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

Elton, E. J., Gruber, M. J., & Blake, C. R. (1995). Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance, 50*(4), 1229-1256.

Elton, E. J., Gruber, M. J., Agrawal, D., & Mann, C. (2001). Explaining the rate spread on corporate bonds. *The Journal of Finance, 56*(1), 247-277.

European Sustainable Investment Forum – EUROSIF. (2016). *European SRI Study 2016.* Retrieved July 28, 2017, from http://www.eurosif.org

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics, 49*(3), 283-306.

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics, 25*, 23-49.

Fama, E. F., & French, K. R. (1992). The cross-section of expected returns. *Journal of Finance, 47*, 426–465.

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

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance, 63*(4), 1653–1678.

Ferson, W. E., Sarkissian, S., & Simin, T. (2003). Is stock return predictability spurious? *Journal of Investment Management, 1*(3), 1-10.

Ferson, W. E., Sarkissian, S., & Simin, T. (2008). Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression. *Journal of Financial and Quantitative Analysis, 43*(2), 331-353.

Ferson, W., & Qian, M. (2004). Conditional performance evaluation, revisited. *The Research Foundation of CFA Institute*.

Ferson, W., & Schadt, R. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance, 51*, 425–462.

Filbeck, G., Gorman, R., & Zhao, X. (2009). The "Best Corporate Citizens": Are they good for their shareholders? *The Financial Review, 44*, 239–262.

Freeman, R. (1984). *Strategic Management: A Stakeholder Approach.* Boston: Pitman.

Friedman, M. (1962). *Capitalism and Freedom.* Chicago: The University of Chicago Press.

Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Return and risk in socially responsible investment. *Journal of Banking & Finance, 32*, 2646–2654.

Ge, W., & Liu, M. (2015). Corporate social responsibility and the cost of corporate bonds. *Journal of Accounting and Public Policy, 34*(6), 597-624.

Gil-Bazo, J., Ruiz-Verdu, P., & Santos, A. A. (2010). The performance of socially responsible mutual funds: The role of fees and management companies. *Journal of Business Ethics, 94*(2), 243–263.

Godfrey, L. G. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica, 46*(6), 1293-1301.

Goldreyer, E. F., & Diltz, J. D. (1999). The performance of socially responsible mutual funds: Incorporating sociopolitical information in portfolio selection. *Managerial Finance, 25*(1), 23-36.

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics, 118*, 107-155.

Gonenc, H., & Scholtens, B. (2017). Environmental and financial performance of fossil fuel firms: A closer inspection of their interaction. *Ecological Economics, 132*, 307–328.

Gregory, A., & Whittaker, J. (2007). Performance and performance persistence of ethical unit trusts in the UK. *Journal of Business Finance & Accounting, 34*(7/8), 1327-1344.

Halbritter, G., & Dorfleitner, G. (2015). The wages of social responsibility — Where are they? A critical review of ESG investing. *Review of Financial Economics, 26*, 25-35.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*, 357—384.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal, 49*(6), 62-66.

Henke, H.-M. (2016). The effect of social screening on bond mutual fund performance. *Journal of Banking & Finance, 67*, 69-84.

Hoepner, A. G., & Nilsson, M. A. (2017). No news is good news: Corporate social responsibility ratings and fixed income portfolios. [Working Paper] SSRN eLibrary.

Hoepner, A., Oikonomou, I., Scholtens, B., & Schröder, M. (2016). The effects of corporate and country sustainability characteristics on the cost of debt: An international investigation. *Journal of Business Finance & Accounting, 43*(1-2), 158-190.

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

Hottinga, J., van Leeuwen, E., & van Ijserlo, J. (2001). Successful factors to select outperforming corporate bonds. *Journal of Portfolio Management, 28*, 88-101.

Ilmanen, A. (1995). Time-varying expected returns in international bond markets. *The Journal of Finance, 50*(2), 481-506.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance, 23*(2), 389-416.

Johnson, R. A., & Greening, D. W. (1999). The effects of corporate governance and institutional ownership types on corporate social performance. *Academy of Management Journal, 42*, 564-576.

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

Kosowski, R. (2006). Do mutual funds perform when it matters most to investors ? US mutual fund performance and risk in recessions and expansions. [Working Paper] SSRN eLibrary.

Kreander, N., Gray, R. H., Power, D. M., & Sinclair, C. D. (2005). Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. *Journal of Business Finance & Accounting, 32*(7/8), 1465-1493.

Leite, C., Cortez, M. C., Silva, F., & Adcock, C. (2018). The performance of socially responsible equity mutual funds: Evidence from Sweden. *Business Ethics: A European Review, 27*(2), 108-126.

Leite, P., & Cortez, M. C. (2015). Performance of European socially responsible funds during market crises: Evidence from France. *International Review of Financial Analysis, 40*, 132-141.

Leite, P., & Cortez, M. C. (2016). The Performance of European socially responsible fixed-income funds. [Working Paper] SSRN eLibrary.

Leite, P., & Cortez, M. C. (2018). The performance of European SRI funds investing in bonds and their comparison to conventional funds. *Investment Analysts Journal*, 47(1), 65-79.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social, capital, trust and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance, 72*(4), 1785-1824.

Liston, D. P., & Soydemir, G. (2010). Faith-based and sin portfolios. *Managerial Finance, 36*(10), 876-885.

Lobe, S., & Walkshäusl, C. (2016). Vice versus virtue investing around the world. *Review of Managerial Science, 10*(2), 303-344.

Lucey, B., & Zhang, Q. (2010). Financial integration and emerging markets capital structure. *Journal of Banking and Finance, 35*(5), 1228-1238.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance, 7*(1), 77-91.

Menz, K.-M. (2010). Corporate social responsibility: Is it rewarded by the corporate bond market? A critical note. *Journal of Business Ethics, 96*(1), 117-134.

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

Mitchell, H., Brown, R., & Easton, S. (2002). Old volatility - ARCH effects in 19th century consol data. *Applied Financial Economics, 12*, 301-307.

Mollet, J. C., & Ziegler, A. (2014). Socially responsible investing and stock performance: New empirical evidence for the US and European stock markets. *Review of Financial Economics, 23*(4), 208–216.

Mollet, J. C., von Arx, U., & Ilić, D. (2013). Strategic sustainability and financial performance: Exploring abnormal returns. *Journal of Business Economics, 83*(6), 577–604.

Moskowitz, M. (1972). Choosing socially responsible stocks. *Business and Society Review*(1), 71-75.

Moskowitz, T. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses: Discussion. *Journal of Finance, 55*(4), 1695-1703.

Munoz, F., Vargas, M., & Marco, I. (2014). Environmental mutual funds: Financial performance and managerial abilities. *Journal of Business Ethics, 124*(4), 551-569.

Newey, W., & West, K. (1987). A simple, positive-definite, heteroskedascity and autocorrelation consistent covariance matrix. *Econometrica, 55*, 703-708.

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

Nollet, J., Filis, G., & Mitrokostas, E. (2016). Corporate social responsibility and financial performance: A non-linear and disaggregated approach. *Economic Modelling, 52*, 400-407.

Oikonomou, I., Brooks, C., & Pavelin, S. (2014). The effects of corporate social performance on the cost of corporate debt and credit ratings. *The Financial Review, 49*, 49–75.

Pagan, A. R., & Sossounov, K. A. (2003). A simple framework for analyzing bull and bear markets. *Journal of Applied Econometrics, 18*(1), 23-46.

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

Revelli, C., & Viviani, J. L. (2015). Financial performance of socially responsible investing (SRI): What have we learned? A meta-analysis. *Business Ethics: A European Review, 24*(2), 158-185.

Reyes, M. G., & Grieb, T. (1998). The external performance of socially-responsible mutual funds. *American Business Review, 16*(1), 1-7.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*(3), 341-360.

Royston, P. (1991). sg3.5: Comment on sg3.4 and an improved D'Agostino test. *Stata Technical Bulletin, 3*, 23-24.

Scholtens, B. (2005). Style and performance of Dutch social responsible investment funds. *Journal of Investing, 14*(1), 63-72.

Schröder, M. (2004). The performance of socialy responsible investments: Investment funds and indices. *Financial Markets and Portfolio Management, 18*(2), 122-142.

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

Schröder, M. (2014). Financial effects of corporate social responsibility: A literature review. *Journal of Sustainable Finance & Investment, 4*(4), 337-350.

Silva, F., & Cortez, M. C. (2016). The performance of US and European green funds in different market conditions. *Journal of Cleaner Production, 135*, 558-566.

Silva, F., Cortez, M. C., & Armada, M. R. (2003). Conditioning information and European bond fund performance. *European Financial Management, 9*(2), 201–230.

Stata Press. (2017). *Stata Press.* Retrieved December 4, 2017, from Stata Base Reference Manual: https://www.stata.com/manuals13/rsktest.pdf

Statman, M. (2006). Socially responsible indexes: Composition, performance, and tracking error. *The Journal of Portfolio Management*, 100-109.

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

Stellner, C., Klein, C., & Zwergel, B. (2015). Corporate social responsibility and Eurozone corporate bonds: The moderating role of country sustainability. *Journal of Banking & Finance, 59*, 538–549.

Taylor, S. J. (2005). *Asset price dynamics, volatility, and prediction.* Princeton: Princeton University Press.

The Business Cycle Dating Committee for the Euro Area of the Centre for Economic Policy Research. (2017, August 7). *Euro Area Business Cycle Dating Committee*. Retrieved October 24, 2017, from Centre for Economic Policy Research: http://cepr.org/content/euro-area-business-cycle-dating-committee

Trinks, P. J., & Scholtens, B. (2017). The opportunity cost of negative screening in socially responsible investing. *Journal of Business Ethics, 140*(2), 193–208.

Utz, S., & Wimmer, M. (2014). Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. *Journal of Asset Management, 15*(1), 72-82.

Van de Velde, E., Vermeir, W., & Corten, F. (2005). Corporate social responsibility and financial performance. *Corporate Governance, 5*(3), 129–138.

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance - Financial performance link. *Strategic Management Journal, 18*(4), 303–319.

Waddock, S. A., Graves, S. B., & Gorski, R. (2000). Performance characteristics of social and traditional investments. *The Journal of Investing, 9*(2), 27-38.

Wang, X. (2010). On time varying mutual fund performance. [Working Paper] Published online:

https://pdfs.semanticscholar.org/e1cd/29480d6cd5b401f1d75ca82fc550559dbe68.pdf.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica, 48*, 817–38.

Wimmer, M. (2013). ESG-persistence in socially responsible mutual funds. *Journal of Management and Sustainability, 3*(1), 9-15.

APPENDICES

Appendix A. Number of bonds and companies for each country and industry

1. Number of bonds and companies for each country

	Bonds	Companies
France	396	61
Germany	144	37
Italy	123	22
Netherlands	76	13
Finland	58	19
Belgium	52	9
Spain	40	15
Austria	36	6
Portugal	9	6
Greece	1	1
Total	935	189

2. Number of bonds and companies for each industry

	Bonds	Companies
Basic Materials	94	24
Consumer Goods	154	26
Consumer Services	108	27
Health Care	37	12
Industrials	217	48
Oil & Gas	46	9
Technology	20	10
Telecommunications	94	11
Utilities	165	22
Total	935	189

Appendix B. Histogram of ESG Scores
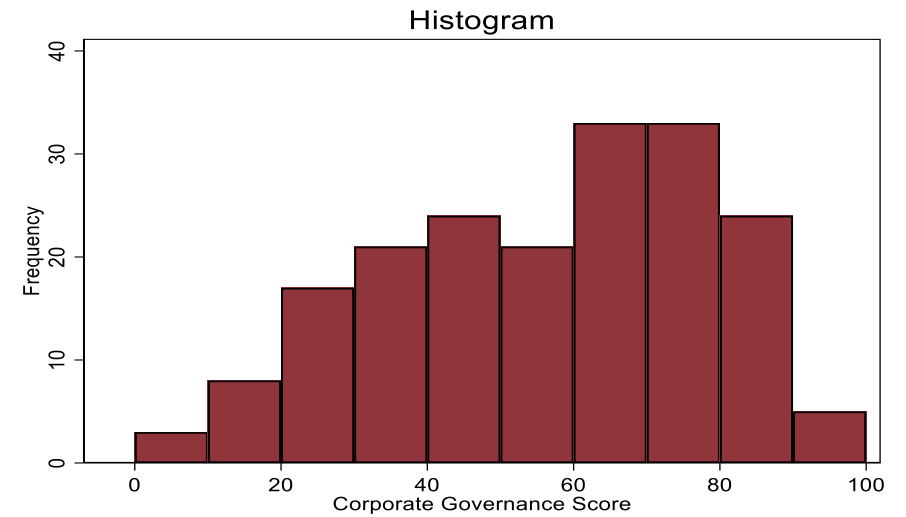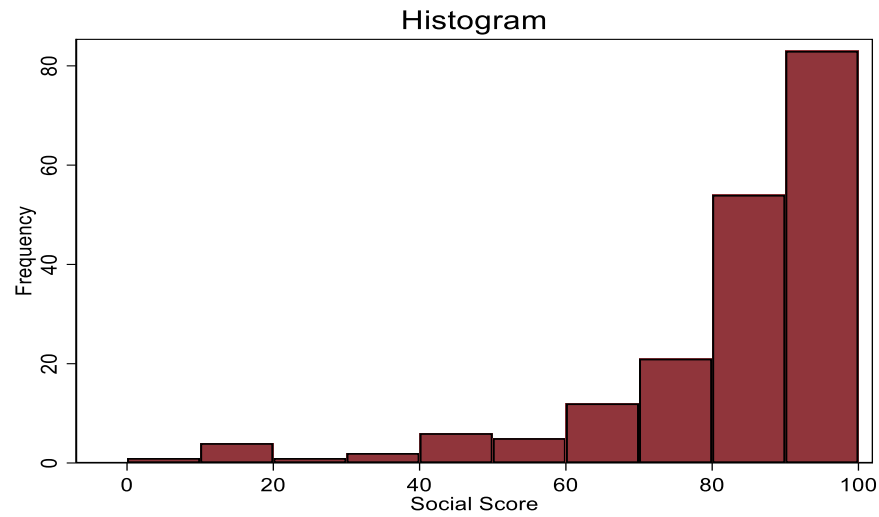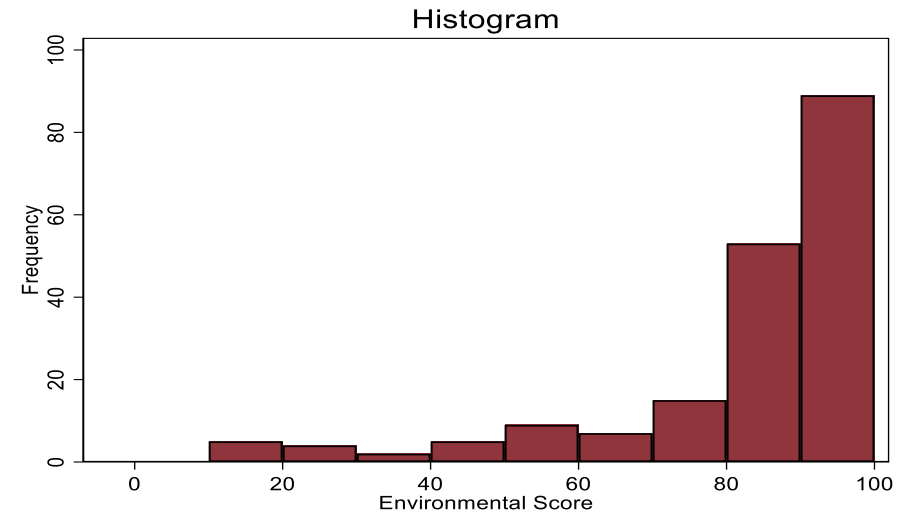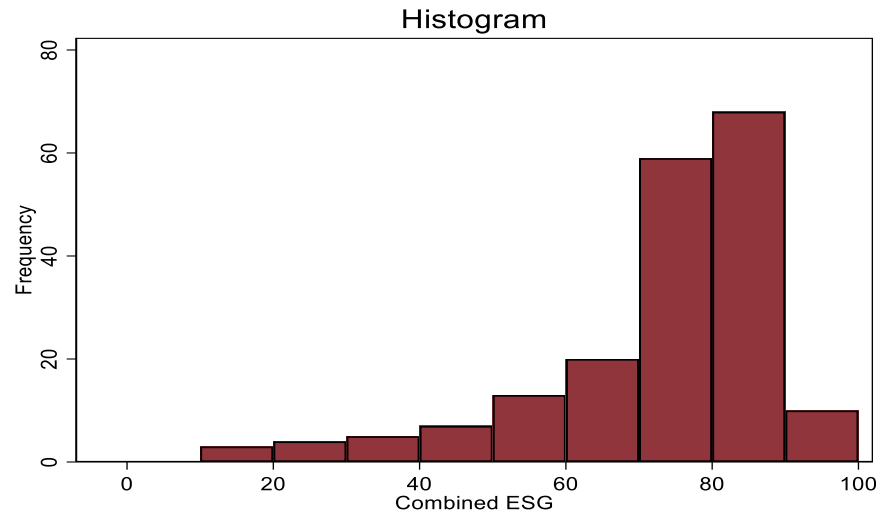
Appendix C. Descriptive statistics of ESG scores on portfolios (positive screening strategy) - alternative cut-offs

This appendix reports descriptive statistics of the ESG scores of the high- and low-rated portfolios. The high-rated (low-rated) portfolios consists of bonds from the 25% and 10% of all companies with the highest (lowest) rating. The sample includes 935 bonds issue by 189 Eurozone companies from the ASSET4 database between 2003 and 2016.

1. Descriptive statistics of ESG scores on portfolios (positive screening strategy) - 25% cut-off

	Environmental Score		Social Score		Governance Score		Combined ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	94.252	68.724	95.925	69.134	84.151	31.572	89.078	59.257
Std. Dev.	0.942	19.524	0.789	17.182	5.778	10.780	2.378	13.992
Median	94.042	76.580	95.853	73.878	84.670	31.777	89.188	64.480
Minimum	92.590	11.066	93.960	8.790	57.470	7.128	74.437	13.308
Maximum	97.380	89.190	98.400	90.103	95.910	52.290	94.637	77.043
Skewness	1.399	-1.425	0.231	-1.502	-2.037	-0.092	-0.932	-1.465
Kurtosis	4.862	4.143	2.521	5.253	9.804	2.438	6.728	4.924
Obs	449	327	482	325	435	324	445	282

2. Descriptive statistics of ESG scores on portfolios (positive screening strategy) - 10% cut-off

	Environmental Score		Social Score		Governance Score		Combined ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	94.761	43.793	96.747	50.202	89.698	19.565	91.408	42.647
Std. Dev.	1.010	19.271	0.572	17.392	3.292	6.464	1.775	13.757
Median	94.385	47.917	96.615	53.405	90.074	21.350	91.728	47.866
Minimum	93.690	11.066	95.640	8.790	64.250	6.500	81.640	13.308
Maximum	97.470	76.580	98.600	83.120	95.910	30.349	94.637	66.937
Skewness	1.436	-0.083	0.902	-0.795	-3.922	-0.271	-2.429	-0.746
Kurtosis	3.772	1.891	3.617	3.122	26.816	1.976	14.045	2.598
Obs	235	114	272	121	240	131	268	100

Appendix D. Credit rating classifications

This appendix reports the classification of credit ratings compiled by Standard & Poor's and the corresponding categories of rating scores. The ordered ranking scale ranges between one for D (lowest rating) and seven for AAA (highest rating). S&P classifies ratings below BBB- as speculative.

S&P debt rating	Assigned rating score	Grade
AAA	7	Investment
AA+	6	Investment
AA	6	Investment
AA-	6	Investment
A+	5	Investment
A	5	Investment
A-	5	Investment
BBB+	4	Investment
BBB	4	Investment
BBB-	4	Investment
BB+	3	Speculative
BB	3	Speculative
BB-	3	Speculative
B+	2	Speculative
B	2	Speculative
B-	2	Speculative
CCC+	1	Speculative
CCC	1	Speculative
CC	1	Speculative
C	1	Speculative
SD	1	Speculative
D	1	Speculative

Appendix E. Exposition of portfolios to investment and speculative grade bonds

This appendix reports the weights of investment and speculative grade bonds in each portfolio. Values are expressed in percentage. The high-rated (low-rated) portfolios consists of bonds from the 50% of all companies with the highest (lowest) rating.

	Investment grade	Speculative grade
Env. Score		
High-rated	86.540	13.460
Low-rated	90.650	9.350
Soc. Score		
High-rated	89.820	10.180
Low-rated	85.710	14.290
Gov. Score		
High-rated	87.550	12.450
Low-rated	89.880	10.120
ESG Score		
High-rated	88.440	11.560
Low-rated	90.780	9.220

Appendix F. Descriptive statistics of portfolios (positive screening strategy) – alternative cut-offs

This appendix reports descriptive statistics on the monthly returns of the high- and low-rated portfolios (25% and 10% cut-offs) as well as the long-short portfolios for each individual and combined ESG score between 2003 and 2016. The high-rated (low-rated) portfolios consists of bonds from the 25% and 10% of all companies with the highest (lowest) rating and the long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The Adj. $\mathcal{X}^2$ is a statistic that is approximately a $\mathcal{X}^2$ distribution with 2 degrees of freedom under the null of normality. The result "." should be interpreted as an absurdly large number and, hence, the data are most certainly not normal. ***, ** and * indicate significance at the 1%, 5%, and 10% level, respectively.

1. Descriptive statistics of portfolios (positive screening strategy) – 25% cut-off

	Mean	Std. Dev.	Median	Minimum	Maximum	Skewness	Kurtosis	Adj. $\mathcal{X}^2$	p-value
Environmental Score									
High-rated	0.424%***	0.828%	0.549%	-2.245%	2.441%	-0.244	2.914	1.790	0.410
Low-rated	0.423%***	0.962%	0.577%	-4.896%	2.768%	-1.064	7.400	36.400	0.000
Long-short	0.001%	0.554%	-0.007%	-1.231%	5.619%	6.259	64.539	.	0.000
Social Score									
High-rated	0.448%***	0.920%	0.542%	-2.304%	3.084%	-0.019	3.102	0.290	0.863
Low-rated	0.436%***	0.925%	0.556%	-4.229%	2.431%	-1.107	6.416	34.170	0.000
Long-short	0.012%	0.590%	-0.047%	-1.009%	5.430%	5.103	45.339	.	0.000
Governance Score									
High-rated	0.412%***	0.961%	0.520%	-2.828%	3.094%	-0.122	3.221	1.140	0.567
Low-rated	0.426%***	0.824%	0.499%	-2.256%	2.511%	-0.253	2.993	1.980	0.373
Long-short	-0.014%	0.497%	-0.008%	-2.575%	1.786%	-0.985	9.641	40.530	0.000
Combined ESG Score									
High-rated	0.450%***	0.974%	0.573%	-2.632%	2.858%	-0.063	3.057	0.290	0.865
Low-rated	0.410%***	0.933%	0.532%	-6.288%	2.460%	-2.205	17.317	.	0.000
Long-short	0.040%	0.702%	-0.023%	-1.870%	7.349%	6.670	71.516	.	0.000

2. Descriptive statistics of portfolios (positive screening strategy) – 10% cut-off

	Mean	Std. Dev.	Median	Minimum	Maximum	Skewness	Kurtosis	X2 value	*p*-value
Environmental Score									
High-rated	0.402%***	0.790%	0.511%	-1.677%	2.729%	0.089	3.082	0.480	0.789
Low-rated	0.429%***	1.305%	0.622%	-11.688%	2.734%	-4.771	45.480	.	0.000
Long-short	-0.027%	1.050%	-0.077%	-1.738%	11.675%	8.260	93.185	.	0.000
Social Score									
High-rated	0.469%***	0.977%	0.571%	-2.286%	3.662%	0.068	3.379	1.520	0.467
Low-rated	0.432%***	0.823%	0.524%	-1.893%	2.461%	-0.312	3.299	3.870	0.145
Long-short	0.037%	0.551%	0.004%	-1.493%	2.568%	0.719	6.080	23.870	0.000
Governance Score									
High-rated	0.403%***	0.950%	0.465%	-2.654%	2.527%	-0.171	3.070	1.090	0.579
Low-rated	0.402%***	0.846%	0.468%	-3.852%	2.573%	-0.773	6.268	25.750	0.000
Long-short	0.000%	0.749%	0.020%	-2.894%	4.429%	0.501	11.474	34.060	0.000
Combined ESG Score									
High-rated	0.411%***	0.995%	0.487%	-3.148%	3.405%	-0.036	3.672	3.030	0.220
Low-rated	0.387%***	1.516%	0.580%	-15.580%	3.293%	-6.992	74.527	.	0.000
Long-short	0.024%	1.322%	-0.051%	-2.501%	15.055%	8.806	101.074	.	0.000

Appendix G. Correlations between factors and information variables

This appendix reports the correlations between the monthly returns of the factors (bond market index, default factor, option factor, equity factor), the risk-free rate (R_f) and the public information variables Term Spread (TS) and Inverse Relative Wealth transformed (i.e., after the stochastic detrending and mean-zero procedures).

1. Correlations between factors

	Market	Default	Option	Equity	R_f
Market	1.000				
Default	0.181	1.000			
Option	-0.306	0.333	1.000		
Equity	0.171	0.647	-0.060	1.000	
Rf	0.097	0.044	0.088	-0.089	1.000

2. Correlations between information variables

	TS (transformed)	IRW (transformed)
TS (transformed)	1.000	
IRW (transformed)	-0.339	1.000

Appendix H. Regression results for the portfolios using the multi-factor model with a broader bond market index (positive screening strategy)

This appendix presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the multi-factor model regressions with the iBoxx € Overall (equation 5). *Bond* corresponds to the monthly excess returns of the iBoxx € Overall index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The high-rated (low-rated) portfolios include bonds from the 50% of all companies with the highest (lowest) rating for each individual and combined ESG score (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ∗∗∗, ∗∗, and ∗ indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary.

	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$
Env. Score						
High-rated	0.036	0.938***	0.094***	0.315***	-0.004	0.786
	(1.08)	(21.87)	(5.79)	(4.66)	(-0.34)	
Low-rated	0.038	0.953***	0.033**	0.397***	0.013	0.759
	(1.08)	(21.06)	(2.24)	(6.48)	(1.30)	
Long-short	-0.002	-0.015	0.060***	-0.081*	-0.017**	0.151
	(-0.07)	(-0.42)	(5.34)	(-1.74)	(-2.28)	
Soc. Score						
High-rated	0.052	0.933***	0.074***	0.312***	-0.004	0.766
	(1.46)	(18.46)	(4.56)	(4.45)	(-0.35)	
Low-rated	0.050	0.914***	0.063***	0.395***	0.014	0.744
	(1.40)	(19.98)	(4.20)	(6.37)	(1.43)	
Long-short	0.002	0.019	0.011	-0.082**	-0.019**	0.039
	(0.07)	(0.43)	(0.64)	(-1.97)	(-2.14)	
Gov. Score						
High-rated	0.038	0.945***	0.084***	0.311***	0.004	0.758
	(1.07)	(21.00)	(4.89)	(4.53)	(0.25)	
Low-rated	0.031	0.944***	0.044***	0.406***	0.000	0.827
	(1.12)	(24.17)	(3.94)	(6.84)	(0.01)	
Long-short	0.007	0.001	0.040***	-0.096**	0.003	0.146
	(0.31)	(0.02)	(3.14)	(-2.05)	(0.30)	
ESG Score						
High-rated	0.044	0.947***	0.090***	0.325***	-0.003	0.781
	(1.31)	(21.62)	(5.67)	(4.83)	(-0.27)	
Low-rated	0.028	0.928***	0.029*	0.390***	0.015	0.747
	(0.78)	(20.39)	(1.93)	(6.33)	(1.50)	
Long-short	0.017	0.018	0.061***	-0.066	-0.018**	0.132
	(0.57)	(0.50)	(5.07)	(-1.32)	(-2.29)	

Appendix I. The performance of portfolios formed on individual and combined ESG scores without outliers (positive screening strategy) - 50%, 25% and 10% cut-off

This appendix presents estimates of monthly abnormal returns (alphas expressed in percentage), factor loadings, and the adjusted R^2 obtained from the conditional multi-factor model regressions (equation 10). *Bond* corresponds to the monthly excess returns of the BofA Merrill Lynch Euro Non-Financial Corp Euro Domicile index. *Default* is a default spread variable, computed the difference in returns between the BofA Merrill Lynch € High-Yield TR index and the iBoxx € Sovereign TR index. The *Option* variable is computed as the difference in return between the BofA Merrill Lynch € Asset-Backed and Mortgage-Backed Securities TR index and the iBoxx € Sovereign TR index. *Equity* corresponds to the monthly excess returns of the FTSE AW Eurozone TR index. Excess returns were computed using the 1-month Euribor. The predetermined information variables are the term spread (TS), the Inverse Relative Wealth (IRW) and a dummy variable for the month of January (JD). The high-rated (low-rated) portfolios include bonds from the 50%, 25% and 10% (panel A, B and C, respectively) of all companies with the highest (lowest) rating for each individual and combined ESG scores (positive screening strategy). The long-short portfolio is a trading strategy going long in the high-rated and short in the low-rated portfolio. The portfolios are value-weighted and rebalanced monthly. The observation period spans the period from 2003 to 2016. ∗∗∗, ∗∗, and ∗ indicate significance at the 1%, 5%, and 10% level, respectively. The values of the t-statistic are presented in parenthesis. Standard errors are computed using the Newey-West (1987) method or the White (1980) variance–covariance matrix, whenever necessary. W_1, W_2 and W_3 correspond to the probability values of the $\mathcal{X}^2$ statistic of the Wald test on the hypothesis that the coefficients of the conditional alphas, conditional betas and conditional alphas and betas, respectively, are jointly equal to zero.

PANEL A	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	Adj. R^2	W_1	W_2	W_3
Env. Score									
High-rated	0.060**	0.848***	-0.001	0.071**	0.007	0.899	0.010	0.486	0.008
	(2.03)	(25.08)	(-0.08)	(2.57)	(1.03)				
Low-rated	0.049**	0.891***	-0.037***	0.084***	0.016***	0.929	0.014	0.002	0.001
	(2.37)	(37.28)	(-3.46)	(3.08)	(3.13)				
Long-short	0.011	-0.043	0.037***	-0.012	-0.009	0.255	0.134	0.000	0.000
	(0.42)	(-1.53)	(3.35)	(0.33)	(-1.52)				
Soc. Score									
High-rated	0.062*	0.875***	-0.017	0.075***	0.009	0.910	0.003	0.031	0.008
	(1.95)	(27.50)	(-1.53)	(3.38)	(1.51)				
Low-rated	0.050**	0.848***	-0.009	0.085***	0.014**	0.908	0.161	0.003	0.002
	(2.13)	(31.88)	(-0.66)	(2.93)	(2.28)				
Long-short	0.012	0.027	-0.008	-0.009	-0.005	0.195	0.167	0.000	0.000
	(0.38)	(1.02)	(-0.60)	(-0.44)	(-0.83)				
Gov. Score									
High-rated	0.061**	0.873***	0.001	0.054	0.010	0.900	0.062	0.390	0.052
	(2.11)	(29.01)	(0.03)	(1.54)	(1.18)				
Low-rated	0.041*	0.855***	-0.040***	0.103***	0.010	0.923	0.013	0.000	0.000
	(1.78)	(35.44)	(-2.60)	(3.20)	(1.37)				
Long-short	0.021	0.019	0.040**	-0.049	0.000	0.222	0.181	0.000	0.000
	(0.73)	(0.84)	(2.11)	(-1.06)	(-0.04)				
ESG Score									
High-rated	0.062**	0.867***	0.003	0.067***	0.006	0.903	0.007	0.454	0.000
	(2.17)	(26.01)	(0.24)	(2.66)	(0.96)				
Low-rated	0.045**	0.857***	-0.050***	0.098***	0.020***	0.924	0.023	0.000	0.000
	(2.26)	(35.67)	(-4.97)	(3.35)	(4.00)				
Long-short	0.018	0.010	0.053***	-0.031	-0.014*	0.316	0.130	0.001	0.000
	(0.81)	(0.41)	(4.30)	(-0.87)	(-1.89)				

PANEL B	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.030	0.824***	-0.008	0.104***	0.009	0.898	0.076	0.002	0.000
	(1.13)	(25.94)	(-0.51)	(3.14)	(1.24)				
Low-rated	0.052**	0.868***	-0.035**	0.062**	0.017**	0.884	0.095	0.000	0.000
	(2.43)	(32.74)	(-2.59)	(2.14)	(2.49)				
Long-short	-0.023	-0.044	0.027**	0.043	-0.008	0.198	0.001	0.000	0.000
	(-0.98)	(-1.219)	(2.18)	(1.25)	(-0.94)				
Soc. Score									
High-rated	0.080*	0.877***	-0.025*	0.071**	0.011	0.874	0.008	0.062	0.024
	(1.90)	(25.79)	(-1.87)	(2.15)	(1.45)				
Low-rated	0.069**	0.792***	-0.014	0.056	0.021**	0.855	0.826	0.057	0.098
	(2.31)	(25.68)	(-0.84)	(1.31)	(2.39)				
Long-short	0.011	0.085*	-0.012	0.015	-0.010	0.261	0.033	0.000	0.000
	(0.24)	(1.94)	(-0.64)	(0.31)	(-1.05)				
Gov. Score									
High-rated	0.024	0.865***	0.008	0.015	0.009	0.867	0.393	0.801	0.353
	(0.74)	(25.50)	(0.41)	(0.34)	(0.95)				
Low-rated	0.063**	0.825***	-0.057***	0.146***	0.017	0.866	0.174	0.130	0.071
	(2.19)	(26.17)	(-2.81)	(3.10)	(1.64)				
Long-short	-0.039	0.040	0.064***	-0.132***	-0.008	0.326	0.783	0.008	0.014
	(-1.17)	(1.14)	(3.52)	(-2.76)	(-0.77)				
ESG Score									
High-rated	0.070**	0.883***	0.011	0.036	0.006	0.871	0.088	0.495	0.091
	(2.04)	(25.32)	(0.54)	(0.77)	(0.56)				
Low-rated	0.098***	0.741***	-0.044**	0.118***	0.026***	0.812	0.071	0.006	0.013
	(3.13)	(22.92)	(-2.56)	(2.64)	(2.87)				
Long-short	-0.028	0.142***	0.054***	-0.082**	-0.020**	0.383	0.175	0.040	0.013
	(-0.97)	(4.74)	(3.43)	(-1.99)	(-2.41)				

PANEL C	α	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W_1	W_2	W_3
Env. Score									
High-rated	0.041	0.760***	-0.002	0.119**	0.011	0.827	0.129	0.020	0.006
	(1.42)	(20.55)	(-0.12)	(2.59)	(1.06)				
Low-rated	0.078*	0.792***	-0.062***	0.068	0.034***	0.743	0.073	0.045	0.041
	(1.84)	(18.02)	(-2.68)	(1.12)	(2.73)				
Long-short	-0.037	-0.032	0.060**	0.051	-0.023*	0.136	0.003	0.004	0.002
	(-0.86)	(-0.71)	(2.54)	(0.83)	(-1.83)				
Soc. Score									
High-rated	0.094*	0.908***	-0.010	0.067	0.010	0.822	0.027	0.057	0.031
	(1.82)	(22.88)	(-0.48)	(1.55)	(0.98)				
Low-rated	0.059	0.795***	-0.031	0.132**	0.014	0.731	0.092	0.321	0.167
	(1.40)	(15.08)	(-1.12)	(2.30)	(1.05)				
Long-short	0.035	0.113**	0.021	-0.065	-0.004	0.152	0.049	0.018	0.002
	(0.68)	(2.27)	(0.78)	(-1.02)	(-0.27)				
Gov. Score									
High-rated	0.007	0.867***	0.004	-0.007	0.002	0.815	0.506	0.389	0.332
	(0.16)	(20.55)	(0.15)	(-0.15)	(0.19)				
Low-rated	0.083**	0.677***	-0.029	0.203***	0.031***	0.710	0.015	0.020	0.005
	(2.24)	(17.58)	(-1.41)	(3.81)	(2.83)				
Long-short	-0.077*	0.190***	0.032	-0.210***	-0.029**	0.269	0.348	0.006	0.006
	(-1.66)	(3.97)	(1.27)	(-3.18)	(-2.16)				
ESG Score									
High-rated	0.011	0.885***	-0.001	-0.018	0.009	0.842	0.349	0.398	0.208
	(0.28)	(22.46)	(-0.06)	(-0.36)	(0.83)				
Low-rated	0.100**	0.778***	-0.085***	0.127***	0.040***	0.708	0.025	0.000	0.000
	(2.49)	(15.47)	(-3.85)	(3.12)	(3.72)				
Long-short	-0.090*	0.107**	0.084***	-0.145**	-0.031**	0.237	0.083	0.028	0.001
	(-1.97)	(2.04)	(3.06)	(-2.24)	(-2.37)				