Universidade do Minho
Escola de Economia e Gestão

Liliana Garcia Gestosa

The performance of socially responsible corporate bond portfolios: empirical evidence for the US market

July 2020

The performance of socially responsible corporate bond portfolios: empirical evidende for the US market

Liliana Garcia Getsosa

UMinho | 2020

Universidade do Minho
Escola de Economia e Gestão

Liliana Garcia Gestosa

The performance of socially responsible corporate bond portfolios: empirical evidence for the US market

Master dissertation
Master's in Finance

Accomplished under the supervision of:
Professor Ph.D. Florinda Cerejeira Campos Silva

July 2020

Acknowledgements

To my grandmother.

I am very proud to finish my dissertation, which represents the end of a chapter. Thus, I want to thank everyone who helped me during this year.

First, I want to thank my supervisor, Professor Florinda Silva, for all the advice, assistance, and for always encouraging me to critically analyze every aspect of my dissertation. Thank you for the patience, the meetings, and all the help throughout this year.

Second, I want to thank my family for always believing in me and supporting my goals. Without you, I wouldn't have been able to accomplish any of this. To my mother, father, and brother, thank you for being my role models and always being by my side.

I also want to acknowledge my friends for all their support, all the days spent working in the library, and the times we made plans just to have fun and relax. You were the best partners I could have asked for during these five years in Braga. Thank you for everything!

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

Resumo

Esta dissertação tem como objetivo avaliar o desempenho financeiro de carteiras de obrigações socialmente responsáveis de empresas dos EUA, entre 2003 e 2018. As classificações de responsabilidade social corporativa são extraídas da *Thomson Reuters ESG Database*, resultando numa amostra composta 8670 obrigações emitidas por 851 empresas. As carteiras são ponderadas pelo valor de mercado das obrigações e construídas usando as abordagens positiva e *best-in-class*, para avaliações individuais e agregadas de responsabilidade social. Para além disso, são construídas carteiras de classificação alta, baixa e de diferença, com rebalanceamento anual. Relativamente à avaliação de desempenho das carteiras, são utilizados modelos multi-fator não condicionais e condicionais. Os resultados globais sugerem que não existem diferenças entre a performance financeira de obrigações emitidas por empresas com *ratings* elevados e *ratings* baixos. Ainda que existam algumas alterações ao nível dos alfas de carteiras com *ratings* elevados e *ratings* baixos, as principais conclusões para a abordagem positiva são robustas ao uso de taxas alternativas para a construção das carteiras, esquemas de ponderação e à exclusão do setor financeiro. Relativamente à estratégia *best-in-class*, existe evidência de *performance* superior de carteiras com avaliações baixas baseadas nas avaliações sociais e agregadas, relativamente a carteiras com avaliações elevadas. No entanto, esta evidência desaparece com a construção de carteiras extremas. Por último, a análise da *performance* ao longo do tempo sugere evidências de rendibilidades anormais apenas a partir de 2007, não existindo evidência a favor das hipóteses *shunned-stock* e *errors-in-expectations*.

Palavras-chave: Avaliação de desempenho financeiro, carteiras de obrigações, ESG, investimento socialmente responsável, obrigações.

Abstract

This dissertation aims to assess the financial performance of US corporate bond portfolios formed based on socially responsible criteria between 2003 and 2018. Corporate social responsibility ratings are retrieved from the Thomson Reuters ESG Database, resulting in a dataset of 8670 bonds issued by 851 companies. Value-weighted portfolios are constructed using the positive and best-in-class screening strategies, for individual and aggregate CSR scores. Furthermore, high-rated, low-rated, and difference portfolios are built, with annual rebalancing. Regarding performance evaluation, unconditional and conditional multi-factor models are used. Overall results suggest that there are no differences between the performance of bonds issued by high-rated and low-rated companies. Although there are some changes in the alphas from high- and low-rated portfolios, the main conclusions reached for the positive approach remain robust to alternative cut-offs, weighting schemes, and the exclusion of the financial sector. Regarding the best-in-class strategy, low-rated portfolios based on social and ESG combined ratings outperform high-rated portfolios, however, this outperformance fades when extreme cut-offs are used. Finally, the performance analysis over time provides evidence of abnormal returns for high- and low-rated portfolios only after 2007, and there is no evidence in favor of the shunned-stock and errors-in-expectations hypotheses.

Key-words: bonds, corporate bond portfolios, ESG, performance evaluation, socially responsible investment.

Index

List of tables

1. Introduction

Recently, socially responsible investing (SRI) has attracted considerable attention from investors and scholars worldwide. The increased interest in corporate social responsibility (CSR) is linked to the expansion of this market sector in the past few years. Moreover, issues faced by firms such as global warming, the scarcity of resources, or social tensions arising from wealth disparity may be linked to the growth of incorporating extra-financial health criteria into the investment decision (Menz, 2010).

According to the European Sustainable Investment Forum (EUROSIF, 2016, p. 9), "Sustainable and Responsible Investment is a long-term oriented investment approach, which integrates Environmental, Social and Governance (ESG) factors in the research, analysis, and selection 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 behavior of companies." There has been an ongoing debate on the impact of socially responsible investments on financial performance, which has been the motivation of several recent studies. Furthermore, the US SIF Foundation's 2018 Report on US Sustainable, Responsible and Impact Investing Trends (2018) states that at the end of 2017, from the total existing $46.6 trillion US professionally managed assets, SRI investments amounted to a total of $12 trillion. This is consistent with a substantial increase in SRI, as in 2016 only $8.098 trillion were ESG-related investments.

Motivated by the growth of SRI, several academics have addressed this matter in their studies to assess the impact on the financial performance of stock and bond portfolios of including CSR criteria into the investment decision. Even though this investment approach was initially developed to equity selection processes, a growing number of investors have been applying it to fixed-income securities, as well. Hereupon, existent literature exploring the financial performance of SRI is far more extensive regarding equity portfolios than fixed-income portfolios.

Thus, prior literature mostly aims to analyze whether SRI adds value to investors, compared to conventional investment strategies. Furthermore, the majority of papers focus on the financial performance of SRI equity mutual funds. Evidence frequently shows no significant differences in the performance of SRI equity mutual funds, compared to their conventional peers, yet, findings are mixed (Bello, 2005; Scholtens, 2005). However, studies focusing on mutual funds present some limitations. The financial performance of mutual funds is highly dependent on the fund manager's skills (Kempf & Osthoff, 2007). Moreover, fund managers tend to change fund holdings over time, thus the SRI label attributed to a fund does not guarantee it engages in socially responsible practices (Auer & Schuhmacher, 2016). Finally, fund returns include not only financial returns from the underlying securities but also management fees (Auer & Schuhmacher, 2016). To overcome these drawbacks from mutual fund performance evaluation studies, another line of research emerged, focusing on synthetic SRI stock portfolios. Again, results vary, however, most studies support that SRI stock portfolios outperform conventional portfolios (Kempf & Osthoff, 2007; Derwall, Guenster, Bauer & Koedijk, 2005; Mollet & Ziegler, 2014). Conversely, some research finds no significant performance differences between both types of portfolios (Halbritter & Dorfleitner, 2015; Auer & Schuhmacher, 2016).

However, very few studies extend this analysis to the fixed-income area. Unlike equity instruments, bonds are affected by less non-diversifiable risk factors, allowing managers to eliminate most of the firm-specific risk associated with bonds through diversification. This exploitation of diversifiable risk allows investors to achieve abnormal returns (Derwall & Koedijk, 2009), thus confirming the importance of conducting such performance evaluation studies. Although some literature addresses the performance of SRI bond funds, to the best of my knowledge only three papers focus on evaluating the effect of companies' ESG ratings on the financial performance of the bonds issued by them. While Hoepner and Nilsson (2017) and Polbennikov, Desclée, Dynkin, and Maitra (2016) target the US market, Pereira, Cortez, and Silva (2019) focus on the European market.

Hence, this dissertation aims to evaluate the performance of SRI US corporate bond portfolios, to assess if investing according to the ESG ratings attributed to bond-issuing companies yields abnormal returns. It contributes to the existing literature by filling the gap in SRI bond portfolio performance evaluation studies, by extending the research conducted by Hoepner and Nilsson (2017) to a conditional multi-factor model, allowing both risk and performance to change over time with the state of the economy. To the best of my knowledge, conditional models have not been applied in any studies regarding the US market. Furthermore, the dataset used in this study is more comprehensive than the ones used in previous literature, as it includes the financial sector, as well as bonds with nonstandard characteristics.

The performance evaluation is conducted using a survivorship bias-free dataset, composed of 8760 bonds issued by 851 US companies, between 2003 and 2018. Environmental, Social, and Governance (ESG) ratings of firms are provided by several agencies like Kinder, Lydenberg, and Domini (KLD), Vigeo, or Thomson Reuters ESG. Due to its advantages relative to other databases, namely its reporting consistency and reliability, ESG scores for this dissertation are collected from the Thomson Reuters ESG Database. Furthermore, data concerning bond returns is collected using Thomson Reuters Datastream.

This dissertation is divided into six sections. Following the introduction, the second section reviews the previous literature on the performance evaluation of socially responsible investments. The third section presents the methodology used to assess portfolio performance. Next, the fourth section describes the collected dataset and the constructed portfolios based on CSR ratings, considering both aggregated and individual environmental, social and governance dimensions. The fifth section discusses the empirical results of this study. Finally, the sixth section presents the most relevant conclusions from this dissertation, as well as its limitations and possible approaches for future research.

2. Literature Review

This section presents a theoretical overview of the existing literature regarding the effects of socially responsible investment on financial performance. SRI has emerged in recent years, becoming a large segment of investment in financial markets all over the world. Therefore, the results of several previous studies are discussed to present the state-of-the-art of the field, mainly concerning the performance of socially responsible stock and bond portfolios.

There are several theories about the impact on financial performance of incorporating ethical screenings into the investment decision, which can be divided into three main hypotheses. The first hypothesis argues in favor of the underperformance of socially screened portfolios and stems directly from modern portfolio theory (Markowitz, 1952), which supports that social screening reduces portfolio performance because it limits the investor's possibilities regarding investing (Rudd, 1981). Furthermore, social screening carries extra costs necessary for monitoring and collecting information, which may also contribute to the underperformance of socially responsible portfolios (Cortez, Silva & Areal, 2009). Moreover, investors who are driven by values usually evade investing in controversial activities such as alcohol, weapons, and tobacco, which are more profitable (Derwall, Koedijk & Ter Horst, 2011; Hong & Kacperczyk, 2009). Since these controversial sectors are usually associated with abnormal returns, investors cannot take advantage of them when they divert their investments from certain activities.

The second hypothesis supports social screening and derives from the stakeholder theory presented by Freeman (1984). According to this hypothesis, firms that take into consideration all stakeholder's interests when investing may have higher gains in the future through increased shareholder value (Jensen, 2001). Moreover, Porter and Kramer (2006) argue that companies may gain a comparative advantage by including socially responsible practices into their activity. Furthermore, social screening may also be an indicator of good management and corporate governance (Waddock & Graves, 1997). Thus, companies engaging in these practices may perform better when compared to companies that do not adopt CSR strategies, while also displaying less default risk (Hoepner, Oikonomou, Scholtens & Schröder, 2016).

Finally, the "no effect" hypothesis states that there is no difference in expected returns between stocks issued by socially responsible firms and their conventional peers. Such may be true when a firm's socially responsible practices are costless and not translating into the creation of financial value. The neutral effect of social screening can occur, as well, when the benefits from socially responsible practices are offset by its costs, thus not influencing firm profitability. Furthermore, when investors overestimate the benefits of CSR or underestimate its costs, the net effect on company gains may also be neutral (Statman & Glushkov, 2009).

Social screens are largely applied in portfolio construction and consist of considering non-financial criteria when investing, as to include or exclude certain securities from investment portfolios. Typically, three types of screening strategies are used, namely the positive (or qualitative), negative (or exclusionary), and best-in-class approaches. Positive social screening concentrates on the inclusion of companies with positive social performance, rather than the exclusion of companies with low CSR ratings, consistent with a negative screening strategy. According to Schwartz (2003), some common negative screening strategies exclude firms associated with controversial businesses, such as tobacco or alcohol, from their pool of investments. Conversely, the positive approach includes some firms specifically due to the activities in which they are involved. Lastly, the best-in-class approach is similar to the positive screening strategy, but the resulting portfolio is balanced across industries (Kempf & Osthoff, 2007).

Most of the recent literature on SRI investigates whether this type of investment approach influences the market and the financial performance of stock and bond portfolios. The applied methodologies are usually based on comparing the financial performance of portfolios with high sustainability scores on one or more ESG dimensions with lower score portfolios. Furthermore, financial performance is usually estimated using risk-adjusted measures such as the Fama and French (2015) five-factor alphas or the Carhart (1997) four-factor alphas and a wide range of the literature focuses on the US market.

2.1 Socially responsible investing in equity portfolios

The financial performance of socially responsible investments has received much attention in the last decades. The study of SRI initiates with stocks, in 1972 (Moskowitz, 1972), evolving to the performance assessment of SRI indexes (Schröder, 2007), and is followed by SRI mutual fund studies (Scholtens, 2005). Although most of the literature focuses on the performance of SRI mutual funds, a new line of investigation recently emerged, concerning synthetic SRI stock portfolios (Kempf & Osthoff, 2007). As previously mentioned, existing literature mostly addresses the US market, and the conclusions drawn from these studies range from finding a positive relationship to a neutral link between social and financial performance.

Ortas, Moneva, and Salvador (2014) assess the performance of SRI strategies in the Eurozone, using a modified state-space specification of the Fama and French (1993) single- and three-factor models. Results suggest that investors can engage in SRI strategies without forgoing financial returns, as the return differences between the chosen SRI stock indexes and benchmarks are not significant. However, the authors conclude that the SRI equity indexes show higher systematic risk relative to conventional indexes, which increases with the screening intensity. These results are in line with Schröder (2007), who finds no difference in the performance from SRI stock indexes relative to conventional stock indexes.

Regarding stock mutual funds, Scholtens (2005) finds no evidence of significant differences between the risk-adjusted returns of Dutch SRI stock mutual funds compared to their conventional peers. Likewise, Bauer, Koedijk, and Otten (2005), using a sample of 103 German, UK, and US mutual funds, find no evidence of risk-adjusted return differences between SRI and conventional equity mutual funds, from 1991 until 2001. The Carhart (1997) four-factor model is applied and results are robust to the inclusion of ethical indexes.

The performance evaluation of SRI stock portfolios is a more recent line of investigation, as mentioned. Derwall et al. (2005) study the performance of SRI stock portfolios based on Innovest Strategic Value Advisors' corporate eco-efficiency scores and find that the high-rated portfolio outperforms the less eco-efficient portfolio. Results remain robust regardless of the performance evaluation model used or the presence of transaction costs, for the period between 1995 and 2003.

One of the most recognized studies is by Kempf and Osthoff (2007). Using a dataset of US companies, the authors investigate whether investors can improve their financial results by incorporating social screens into the investment process, from 1992 until 2004. CSR ratings are collected from KLD, and the Carhart (1997) four-factor model is used to evaluate performance. Results show that investors can achieve abnormal returns by engaging in a positive screening strategy consisting of a long position in the high-rated portfolio and a short position in the portfolio with low CSR ratings. Furthermore, the highest alphas derive from the best-in-class investment strategy, especially when top-rated companies are selected. The negative screening approach, however, does not present any significant results. Regarding the performance of specific CSR dimensions, Kempf and Osthoff (2007) find that a positive screening investment strategy yields significant alphas for the employee relations and community dimensions. When it comes to community screening, both the high-rated and long-short portfolios lead to abnormal returns, whereas for employee relations screens only the high-rated portfolio outperforms the benchmark. Findings are robust to alternative portfolio weighting schemes.

Statman and Glushkov (2009) investigate a sample of US firms between 1992 and 2007, using CSR ratings from KLD. According to the authors, a best-in-class investment strategy yields significant abnormal returns. Furthermore, a strategy consisting of excluding stocks from companies associated with controversial businesses results in disadvantages for investors. However, the net effect on performance is neutral, as the advantages of SRI are offset by the loss of shunning controversial stocks. These results contrast with those of Kempf and Osthoff (2007).

Filbeck, Gorman, and Zhao (2009) run an event study in which they analyze the financial performance of US companies included in the *100 Best Corporate Citizens*, for the 2000-2007 period. Results show that the screened portfolio, consisting of stocks issued by firms included in the ranking, consistently outperforms the S&P 500, however, performance is neutral when compared to the matched portfolio. Moreover, an annual rebalancing strategy that eliminates consecutive winners and includes recently listed companies outperforms both the matched portfolio and the S&P 500. Likewise, Brzeszczyński and McIntosh (2013) conduct a study focusing on UK SRI stocks, using the *Global-100 Most Sustainable Corporations in the World* and the Fama and French (1993) three-factor model. Findings from this study are similar to those from Filbeck et al. (2009), suggesting that the portfolio composed of stocks from the *Global-100* list outperforms the benchmarks, for the 2000-2010 period.

Mollet and Ziegler (2014) assess the performance of both the US and European markets, unlike the previously described studies. They use worldwide CSR data from the Zurich Cantonal Bank, over the period from 1998 until 2009, and the Carhart (1997) four-factor model to evaluate the financial performance of firms. Results show no significant abnormal returns for SRI stocks in both European and US markets.

Halbritter and Dorfleitner (2015) analyze ESG based stock portfolios for US firms between 1991 and 2012, collecting data concerning CSR ratings from ASSET4, KLD, and Bloomberg. Findings suggest no significant performance differences between high- and low-rated ESG stock portfolios. Results are consistent for both aggregate and individual CSR dimension scores, regardless of the data provider. Moreover, their findings are robust to different portfolio cut-offs and weightings. In contrast to Kempf and Osthoff (2007), there is no evidence of outperformance of portfolios constructed considering the individual ESG dimensions, for none of the used cut-offs or sources of the ESG scores.

Auer (2016), analyzes the European market from 1991 until 2012 and investigates whether an investment strategy built on SRI adds value to stock portfolios. He finds that investors can earn abnormal returns by engaging in specific screening strategies. Whereas positive social screens may impact portfolio performance negatively, a negative screening strategy yields positive and statistically significant abnormal returns. Furthermore, results show that a negative screening investment strategy consisting of excluding stocks without rating causes meaningful outperformance, compared to a passive benchmark strategy. In relation to specific ESG screens, a negative screening strategy based on the environmental and social pillars does not generate significant abnormal returns, as these portfolios neither destroy nor add value for investors. Nevertheless, results suggest that corporate governance-based portfolios yield significant abnormal returns, under the same conditions. These findings are in line with Halbritter and Dorfleitner (2015) regarding social and environmental portfolios, while contrasting with the evidence of Kempf and Osthoff (2007) for the US market regarding portfolios constructed on social criteria.

Auer and Schuhmacher (2016) evaluate the performance of socially screened stock portfolios across the United States, Asia-Pacific, and Europe markets over the period 2001-2012. Their results suggest that the location and industry of a given SRI investment strategy impact portfolio performance. Hence, for investors focusing on the US and Asia-Pacific region markets, there appears to be no evidence that selecting stock portfolios based on ESG ratings affects earnings, allowing investors to follow an ESG-based investment strategy without sacrificing financial performance. When it comes to the European market, however, investors may pay a price for engaging in SRI strategies for certain industries or ESG criteria. In sum, overall findings provide no evidence supporting the superiority of an investment strategy consisting of high or low-rated ESG stocks comparing to conventional investment approaches.

2.2. Socially responsible investment in fixed-income securities

Previous literature is far more extensive regarding the performance evaluation of stocks, rather than bonds. Furthermore, most of the bond studies concern actively managed fixed-income funds. Blake, Elton, and Gruber (1993) present the first major study focusing on the financial performance of bond mutual funds, between 1979 and 1988. Their investigation is conducted by comparing the performance of actively managed bond funds with several market indexes with similar risk levels such as the Lehman Brothers (LB) government/corporate bond index, the LB mortgage-backed securities index, the Blume/Klein high-yield index, and the LB corporate index. Overall results support the underperformance of bond mutual funds relative to the chosen indexes.

Elton, Gruber, and Blake (1995) develop relative pricing models to evaluate the financial performance of bond mutual funds. The constructed models include bond indexes (as well as an equity index) and unexpected changes in macroeconomic variables such as the expected inflation rate and the real gross national product (GNP). Findings suggest that bond funds yield lower returns than those predicted by the proposed model. In addition, return indexes present the most power in explaining returns.[1] Furthermore, since the factors chosen for their model perform well in explaining bond returns, following investigations such as Derwall and Koedijk (2009) and Henke (2016) use these risk factors.

Silva, Cortez, and Armada (2003) assess the performance of bond funds for the European market. They use a sample of funds from countries such as Portugal, Spain, France, Germany, and the UK, between 1994 and 2000. Overall results suggest that bond funds underperform relative to passive investment strategies, especially funds from Portugal, Spain, Italy, and UK "Gilt" funds. However, German, UK "Corporate", UK "Other Bond", as well as many French funds exhibit neutral financial performance. Findings are robust for both single- and multi-index models.

Another line of research focuses on the financial performance of SRI bond funds, in comparison to their conventional peers. Derwall and Koedijk (2009) present one of the most cited studies in this field, in which they analyze the performance of US SRI bond and balanced mutual funds between 1987 and 2003, using a four-factor model that includes an option factor, a stock market factor, a default spread factor, and a bond market factor. Results show that SRI balanced

[1] Since the performance evaluation model used in this dissertation follows the one developed by Elton et al. (1995), a more detailed description of the factors included in the model is presented in Chapter 3.

funds outperform their conventional peers by about 1,3% per year, whereas SRI fixed-income funds do not perform better than conventional funds.

Henke (2016) investigates the financial performance of SRI bond funds between 2001 and 2014 using a sample of 103 US and Eurozone SRI bond funds. He applies a five-factor model with an aggregate bond market factor, a default factor, an option factor, an equity factor, and a term factor, and finds evidence supporting the outperformance of SRI bond funds relative to their conventional peers, as well as the market. These results contrast with those of Derwall and Koedijk (2009), regarding SRI fixed-income funds.

Leite and Cortez (2018) study the performance of European fixed-income and balanced SRI funds, specifically funds domiciled in France and Germany, from 2002 until 2014, using data from Vigeo. They apply a conditional six-factor model and conclude that SRI fixed-income funds display a better performance than conventional funds. These results are in line with Henke (2016), however, contrast with Derwall and Koedijk (2009). Furthermore, balanced SRI funds do not display statistically significant differences in performance relative to their conventional counterparts. These results are contrasting with the conclusions reached by Derwall and Koedijk (2009), as they find evidence supporting the outperformance of balanced funds compared to their conventional counterparts.

Few studies have been published about the performance of synthetic SRI bond portfolios. Focusing on the US market, Polbennikov et al. (2016) investigate whether ESG ratings affect bond returns, for a sample period between 2007 and 2015. Findings show that investors can achieve abnormal returns by constructing socially responsible bond portfolios based on both individual and aggregate ESG ratings. However, the governance score has the largest influence on abnormal returns for investors, whereas the contribution of the social and environmental pillars is slightly smaller.

Hoepner and Nilsson (2017) extend the analysis of Kempf and Osthoff (2007) to fixed income securities. They investigate synthetic corporate bond portfolios to determine whether the ESG rating of a company influences the performance of the bonds issued by that company. The used sample is composed of 5240 US bonds from 425 corporations, over the period 2001-2014, and a nine-factor model is used to measure portfolio performance. The authors find that bonds issued by companies with no strengths or concerns outperform by 0.89% per year when compared to bonds issued by firms with strengths or concerns, concluding that "no news is good news" when it comes to SRI bond portfolios. These findings contrast with previous literature concerning the performance of SRI stock portfolios (Kempf & Osthoff, 2007; Statman & Glushkov, 2009), and fixed-income portfolios (Polbennikov et al., 2016). Moreover, Hoepner and Nilsson (2017) analyze portfolios based on the individual categories included in the KLD database, namely diversity, community, employee relation, environment, human rights, and product safety. Results suggest significant outperformance of bonds issued by companies with neutral or no CSR scores for the community, product safety, and diversity categories. In contrast to existing empirical evidence (Kempf & Osthoff, 2007; Statman & Glushkov, 2009), the authors do not find evidence of abnormal returns of portfolios based on the human rights and employee relation categories. Furthermore, in relation to the environment category, both high- and low-rated portfolios underperform the benchmark. Finally, findings show that bonds issued by companies with no strengths, concerns, or controversies outperform the market, as well as bonds issued by firms with controversies by 1.12%.

Pereira et al. (2019) investigate the performance of socially screened synthetic bond portfolios between 2003 and 2016, using a dataset of bond issued by 189 Eurozone firms and a conditional four-factor model. Results suggest that investors can adopt an investment strategy based on bonds from firms with high ESG ratings without compromising performance and returns. However, the long-short portfolios do not show any statistical significance. Moreover, findings show that high-rated bond portfolios yield significant positive abnormal returns for the environmental and social dimensions, but not for the governance pillar of ESG.

2.3 Performance of socially responsible investment over time

Another line of research evaluates the performance of SRI over time. Derwall et al. (2011) claim that the SRI movement can be segmented into values-driven and profit-seeking investors. Values-driven investors are not motivated by monetary considerations when adopting an SRI investment strategy, as their investment approach is based on personal and social standards. Profit-seeking investors, in contrast, are motivated by conventional financial objectives. Derwall et al. (2011) propose two hypotheses regarding the performance of socially screened portfolios, namely the shunned-stock hypothesis and the errors-in-expectations hypothesis. The shunned-stock hypothesis considers that values-driven investors avoid controversial stocks, which lowers their demand and causes them to yield higher returns. Conversely, the errors-in-expectations hypothesis argues that SRI stocks can generate abnormal returns when markets do not incorporate the effect of SRI practices on firms expected future cash flows, meaning stock prices will not fully reflect all the value-relevant information. Taking these hypotheses as a starting point, the authors construct two portfolios and evaluate their performance over time, between 1992 and 2008, using CSR scores from KLD. Results are in line with the authors' expectations of prevalent abnormal returns of the shunned-stock portfolio and abnormal returns of the errors-in-expectations portfolio that disappear over time. Findings suggest that the shunned-stock portfolio shows statistically significant annualized abnormal returns ranging between 2,58% and 2,86% throughout the period in analysis. Conversely, the errors-in-expectations portfolio, based on strong employee relations criterion, shows decreasing annualized abnormal returns over time from 5,62% to 2,81%. These findings suggest that, as time goes by, investors improve their knowledge about the effects of CSR on companies' expected cash flows, which should be reflected in their stock valuations.

Bebchuck, Cohen, and Wang (2013), similarly to Derwall et al. (2011), find evidence supporting the errors-in-expectations hypothesis. They investigate the performance of SRI stock portfolios between 1990 and 2008 and results suggest that as investors gradually increase their knowledge about corporate sustainability, the mispricing of stocks disappears over time. In the first subperiod of the study, between 1990 and 1999, SRI stock portfolios show positive abnormal returns, while in the second subperiod, from 2000 until 2008, the returns from these portfolios are statistically insignificant. These findings are consistent with the learning hypothesis, as SRI stock portfolios cease to yield abnormal returns when market participants started to pay more attention to governance issues.

Borgers, Derwall, Koedijk, and Ter Horst (2013), using a sample of US firms over the period 1992-2009 and different subperiods, find evidence that errors in investors' expectations lead to the mispricing of stakeholder information. However, as attention to stakeholder information increases, this outperformance disappears because investors adjust their expectations. Results are consistent with the errors-in-expectations hypothesis, as over the first subperiod (1992-2004) abnormal returns are statistically significant, while during the second subperiod (2004-2009) returns are largely statistically insignificant.

Pereira et al. (2019) find empirical evidence of outperformance of SRI fixed-income portfolios over conventional portfolios in the earlier period of their sample. Moreover, they find that until the start of the financial crisis, investors were able to earn abnormal returns by adopting a long-short strategy in SRI bond portfolios based on the environmental and social scores, while the same does not happen with corporate governance portfolios. However, as markets begin to acknowledge the value of CSR practices, investors should no longer expect to achieve abnormal returns from portfolios based on aggregate ESG ratings.

2.4 Performance of socially responsible investment during crisis periods

In this section, some of the empirical evidence on the performance of SRI portfolios during periods of crisis is discussed. It is relevant to evaluate whether adopting SRI strategies throughout periods of market turmoil allows investors to achieve abnormal returns and outperform conventional investment approaches. Some academics believe that investors pay closer attention to corporate behavior and firm risks during recessions or crisis periods (Hirshleifer, 2008; Statman & Shefrin, 1993). Henke (2016) argues that firms are less exposed to risk when their CSR scores are high. Additionally, if investors examine ESG risk more when the economy is weak, ESG screening should have a sharper positive effect on SRI bond portfolio performance throughout crisis periods. Other authors claim that firms' socially responsible characteristics make them less risky during market turmoil periods (Nofsinger & Varma, 2014)

Lins, Servaes, & Tamayo (2017) analyze the impact of CSR scores on firm performance during the financial crisis between 2008 and 2009, when there was an unanticipated increase in the importance of trust. They find that firms with high CSR ratings perform significantly better than low CSR score firms, during crisis periods. Moreover, the authors argue that companies can engage in CSR activities as a tool to build trust among both stakeholders and investors, as conclusions show that in areas where individuals are more trusting, abnormal returns are higher. Results are robust for a set of control variables that account for risk factors and firm characteristics.

Henke (2016) analyzes the performance of bond portfolios during crisis and non-crisis periods, using bear and bull markets as proxies, respectively. He finds that, during crisis periods, SRI bond funds from both the U.S. and the Eurozone outperform conventional funds by 0,65% and 0,92%, respectively.

Polbennikov et al. (2016) conclude that, during the financial crisis, the governance and environmental dimensions display negative risk-adjusted returns. The social pillar, however, shows positive but volatile returns throughout this period. Hoepner and Nilsson (2017) find that their significant results derive almost fully from the second half of the sample, which is defined by the authors as a "post-crisis period".

3. Methods

This chapter presents the methods applied to conduct the investigation in this dissertation. While the first subsection focuses on the portfolio construction process, the following subsections present unconditional and conditional evaluation models used to assess the financial performance of the bond portfolios.

3.1. Portfolio construction

This research aims to determine whether ESG ratings of bond issuing companies lead investors to achieve abnormal returns when a bond investment strategy is adopted. Following Hoepner and Nilsson (2017), value-weighted portfolios are constructed based on companies' end of year CSR scores from the previous period, as scores are collected at the end of each year[2]. Since ESG ratings do not suffer any significant changes throughout the year, all portfolios are rebalanced annually. Thus, at the beginning of period $\boldsymbol{t}$, portfolios are formed based on companies' end of year social scores from period $\boldsymbol{t-1}$ and held until the end of period $\boldsymbol{t}$.

Portfolios are constructed for each dimension of ESG and for the ESG combined score, using positive and best-in-class screening strategies and a 25% cut-off rate. Portfolios from the positive screening strategy are composed of bonds issued by the 25% top-rated companies, as well as the 25% bottom-rated companies. The best-in-class approach, applied to control for industry differences, as well as overcoming any possible industry biases induced by the positive screening policy (Kempf & Osthoff, 2007), consists of selecting the top 25% performers for each industry, as well as the bottom 25%, and building high and low-rated bond portfolios. Following Hoepner and Nilsson (2017), the Industry Classification Benchmark (ICB) is used to organize the companies included in the dataset into eleven different industries. Moreover, an investment strategy of going long in the high-rated portfolio and short in the low-rated portfolio is adopted, representing a long-short strategy. In this investment strategy, the alpha represents the return differences between the best and worst ESG performers.

[2] Some scholars such as Fama and French (1993) choose to rebalance the portfolios in their studies in June rather than December due to the end of the fiscal year and the accounting information necessary to construct the portfolios, which may be biased. However, as this investigation follows Hoepner and Nilsson (2017), portfolios are rebalanced in December.

Regarding bond data, total return index series are retrieved for each bond included in the dataset using Datastream and discrete returns are calculated as in Equation 1.

$$\boldsymbol{R_{i,t}} = \frac{\boldsymbol{RI_t} - \boldsymbol{RI_{t-1}}}{\boldsymbol{RI_{t-1}}} \tag{1}$$

where $\boldsymbol{R_{i,t}}$ is the discrete rate of return of bond $\boldsymbol{i}$ in month $\boldsymbol{t}$, $\boldsymbol{RI_t}$ is the total return index of a given bond in period $\boldsymbol{t}$, and $\boldsymbol{RI_{t-1}}$ is the total return index of a given bond in period $\boldsymbol{t-1}$.

Since companies may issue more than one bond each year, one time-series of bond returns is created, for each year, with the value-weighted returns of the bonds, to proxy for the return of each company in each year, following Hoepner and Nilsson (2017). Value-weighted industry portfolios are combined into a single portfolio by computing the weighted average of bonds from each industry. The value-weighted returns are calculated using the following equation:

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

where $\boldsymbol{R_{p,t}^{VW}}$ is the rate of return of a value-weighted portfolio in month $\boldsymbol{t}$, $\boldsymbol{MV_{i,t-1}}$ is the lagged market value (at the beginning of the month) for each bond $\boldsymbol{i}$ at month $\boldsymbol{t-1}$, and $\boldsymbol{R_{i,t}}$ is the discrete return of bond $\boldsymbol{i}$ in month $\boldsymbol{t}$.

To evaluate if the results from this study are robust to different weighting schemes, equally-weighted portfolios are also constructed.

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

The equally-weighted return of a portfolio $\boldsymbol{p}$ in month $\boldsymbol{t}$, $\boldsymbol{R_{p,t}^{EW}}$, is given by the summation of the discrete returns of all bonds in portfolio $\boldsymbol{p}$ in month $\boldsymbol{t}$, and $\boldsymbol{N}$ is the total number of bonds in the portfolio.

3.2. Performance evaluation models

3.2.1. Unconditional model

Single-index models such as the Capital Asset Pricing Model (CAPM), although easier to apply and widely used in early literature regarding performance evaluation, have some limitations, as they assume a linear relationship between risk and return. However, extensive empirical evidence shows that several other risk factors are priced and have explanatory power over returns (Banz, 1981; Fama & French, 1992). Therefore, using single-index models to assess financial performance may lead investors to incorrect estimations of underperformance or outperformance of portfolios.

The need to overcome the drawbacks of single-index models motivated the development of multi-factor models. Thus, the unconditional multi-factor model used in this investigation is proposed by Elton et al. (1995). It is a four-factor model with a bond market index factor, a default spread factor, an option factor, and a stock market factor and was applied in previous recognized fixed-income performance evaluation studies such as Derwall and Koedijk (2009), and Henke (2016). The unconditional model is given by the following expression:

$$\boldsymbol{r_{p,t}} = \propto_{\boldsymbol{0p}} + \boldsymbol{\beta_{1p} Bond_t} + \boldsymbol{\beta_{2p} Default_t} + \boldsymbol{\beta_{3p} Option_t} + \boldsymbol{\beta_{4p} Equity_t} + \boldsymbol{\varepsilon_{p,t}} \qquad (4)$$

where $\boldsymbol{r_{p,t}}$ represents the excess returns of portfolio $\boldsymbol{p}$ in month $\boldsymbol{t}$; $\propto_{\boldsymbol{0p}}$ is the unconditional alpha and measures abnormal portfolio returns; $\boldsymbol{Bond_t}$ is a bond market factor that captures the exposure to the bond market, as it designates excess returns from the corporate bond market index. $\boldsymbol{Default_t}$ proxies for changes in economic conditions that may alter the likelihood of bond default and is calculated as the return spread between a high-yield bond index and a government bond index. $\boldsymbol{Option_t}$ is computed as the return difference between a mortgage-backed securities index and a treasury bond index and aims to capture the exposure to both option characteristicrelated returns and securitized debt. $\boldsymbol{Equity_t}$ is a stock market factor that designates excess returns from the equity market index and is included in the model to account for both convertible bonds and bond exposure to the stock market; $\boldsymbol{\varepsilon_{p,t}}$ is the error term.

3.2.2. Conditional model

Unconditional performance evaluation models assume that risk and expected returns remain constant over time, regardless of market conditions and the state of the economy. Therefore, under these circumstances, they may lead to incorrect estimates, due to the neglect of the time-varying characteristics of risk and expected return (Cortez et al., 2009).

Conditional performance evaluation models, conversely, use predetermined public information variables (PIV) as a measure of market conditions, allowing for alphas and betas to vary over time with the state of the economy. This assumption is more consistent with reality and, consequently, controls for the biases that arise from unconditional models. Ferson and Schadt (1996) show that unconditional approaches may confound normal variations in risk and expected returns with market timing and stock picking abilities. Furthermore, as the information contained in these variables is public, it can be used by investors to update their forecasts of expected returns. Authors such as Fama and French (1989) point out the relevance of PIV in forecasting returns of both equity and fixed-income securities.

Ferson and Schadt (1996) propose a conditional model in which the conditional beta is time-varying and a linear function of a vector of predetermined information variables used as a proxy for the state of the economy, $\boldsymbol{Z_{t-1}}$, whereas alpha remains constant. The vector $\boldsymbol{Z_{t-1}}$ represents lagged information that are publicly available at time $\boldsymbol{t-1}$, and are used to forecast returns at time t. The model is given by the following equation:

$$\boldsymbol{r_{p,t} = \propto_p + \beta_{0p} r_{m,t} + \beta'_p (z_{t-1} r_{m,t}) + \varepsilon_{p,t}} \quad (5)$$

where $\boldsymbol{\propto_p}$ is the conditional alpha which is equal to zero if investors only use available public information incorporated in $\boldsymbol{Z_{t-1}}$, consistent with neutral performance.

As mentioned, the conditional beta is a linear function of a vector Z_{t-1}, which is a proxy for market conditions, and is given by the following expression:

$$\boldsymbol{\beta_p(Z_{t-1}) = \beta_{0p} + \beta'_p z_{t-1}} \tag{6}$$

where $\boldsymbol{z_{t-1} = Z_t - E(Z)}$ is a vector representing the deviations of $\boldsymbol{Z_t}$ from the unconditional mean values; $\boldsymbol{\beta_{0P}}$ is the average beta and $\boldsymbol{\beta'_p}$ designates the response of the conditional beta coefficients regarding the PIV, $\boldsymbol{Z_t}$.

The conditional model used in this investigation follows the specification presented by Christopherson, Ferson, and Glassman (1998). It extends the model proposed by Ferson and Schadt (1996), allowing both alpha and beta to be time-varying, therefore, alpha is also a linear function of the vector of predetermined information variables, Z_{t-1}. The model is represented by Equation 7.

$$\boldsymbol{r_{p,t} = \propto_{0p} + A'_p Z_{t-1} + \beta_{0p} r_{m,t} + \beta'_p (Z_{t-1} r_{m,t}) + \varepsilon_{p,t}} \tag{7}$$

where $\boldsymbol{\beta_{0p}}$ is the unconditional beta; $\boldsymbol{\beta'_p}$ is a vector that captures the sensitivity of the conditional beta to deviations of $\boldsymbol{Z_{t-1}}$ from its mean value; $\boldsymbol{A'_p}$ is also a vector and it measures the response of the conditional alpha to the conditioning information variables; $\propto_{\boldsymbol{0p}}$ is the average conditional alpha and $\boldsymbol{\varepsilon_{p,t}}$ is the error term.

The time-varying alpha included in the conditional model from Christopherson et al. (1998) is a linear function of $\boldsymbol{Z_{t-1}}$ and is expressed by Equation 8:

$$\propto_{\boldsymbol{p}} (\boldsymbol{Z_{t-1}}) = \propto_{\boldsymbol{0p}} + \boldsymbol{A'_p Z_{t-1}} \tag{8}$$

where $\propto_{\boldsymbol{p}}$ is the conditional alpha and $\boldsymbol{Z_t - 1}$ is a vector containing public information.

Thus, the conditional multifactor extension applied to the previously presented unconditional model leads to the resulting expression, given by the following regression:

$$\begin{aligned} r_{p,t} = \propto_{0p} + A'_p Z_{t-1} + \beta_{1p} Bond_t + \beta'_{1p}(Z_{t-1} Bond_t) + \beta_{2p} Default_t \\ + \beta'_{2p}(Z_{t-1} Default_t) + \beta_{3p} Option_t \\ + \beta'_{3p}(Z_{t-1} Option_t) + \beta_{4p} Equity_t \\ + \beta'_{4p}(Z_{t-1} Equity_t) + \varepsilon_{p,t} \end{aligned} \tag{9}$$

where β_{1p}, β_{2p}, β_{3p}, and β_{4p} are unconditional betas; β'_{1p}, β'_{2p}, β'_{3p} and β'_{4p} are vectors that capture the sensitivity of conditional betas to deviations of Z_{t-1} from their mean values; A'_p is also a vector and it measures the response of the conditional alpha to the conditioning information variables; $\propto_{0p}$ is the average conditional alpha and $\varepsilon_{p,t}$ is the error term.

Conditional performance evaluation models include lagged information variables as a proxy for the state of the economy, as formerly stated, by considering the information investors have access to when making the investment decision. This approach is consistent with a semi-strong form of market efficiency (Ferson & Schadt, 1996). Recent empirical evidence shows that these variables display relevant explanatory power of bond excess returns (Ilmanen, 1995; Leite & Cortez, 2018; Silva et al., 2003. In this study, a set of three information variables is used, namely the term spread, the short-term rate, and inverse relative wealth (IRW).[3] The choice of these PIV is motivated by previous research showing their usefulness in predicting bond returns. Furthermore, all PIV are lagged to represent available information at the time of the portfolio construction (Ilmanen, 1995; Adcock, Cortez, Armada & Silva, 2012).

The term spread is computed as the difference between the annualized yield of a long-term bond and a short-term bond yield. This variable is used as a proxy for the overall expected bond risk premium and has been used in previous studies by Ilmanen (1995), Adcock et al. (2012), and Leite and Cortez (2018)[4]. The short-term rate is used as a measure for expected inflation and

[3] Some previous studies use a dummy variable for the month of January to account for any seasonality effects in risk and returns. However, the decline in its use in more recent literature motivated its exclusion from this investigation.

[4] Some studies combine the term spread and the real bond yield as information variables to capture bond risk premium as the term spread, alone, unlikely captures all the expected changes in bond returns. However, the real bond yield was not used as an information variable in this study due to its reported high correlation with the term spread, as both PIV are calculated as a yield spread between a long-term bond yield and another yield (Ilmanen, 1995; Silva et al., 2003).

it is proxied by short-term interest rates. It was previously used by Cortez et al. (2009) and Ayadi and Kryzanowski (2011). Finally, the inverse relative wealth is used as a proxy for time-varying risk aversion. It is measured as the ratio between past and current levels of wealth. Past real wealth is proxied by the exponentially weighted average of a stock market index, deflated by the consumer price index (CPI), following Ilmanen (1995) and Leite and Cortez (2018), as shown in Equation 10.

$$\boldsymbol{IRW = \frac{ewaW_{t-1}}{W_t} = \frac{W_{t-1} + coef * W_{t-2} + coef^2 * W_{t-3} + \cdots) * (1 - coef)}{W_t}} \quad (10)$$

where $\boldsymbol{W_t}$ represents the level of real wealth in period $\boldsymbol{t}$, $\boldsymbol{ewaW_{t-1}}$ is the exponentially weighted average of the level of real wealth in the lagged period, $\boldsymbol{t-1}$, and $\boldsymbol{coef}$ is the smoothing coefficient.

4. Data

This section presents the dataset used for the study. First, the data collection process is presented. Second, the dataset is described, as well as each of the CSR rating dimensions. Then, some descriptive statistics of the final dataset, factor returns, and PIV are presented.

4.1. Sample construction

This research focuses on the performance evaluation of socially responsible corporate bond portfolios in the US market, which is one of the largest markets worldwide. This selection is motivated by its high level of transparency and data availability. The initial dataset is composed of the universe of US firms covered by the Thomson Reuters ESG database, from 2003 until 2018[5].

Thomson Reuters Datastream was used to collect all the data on corporate bonds included in the dataset. First, identifiers for all companies covered by the ASSET 4 constituents list were selected, resulting in a preliminary dataset composed of 2521 firms. Following, the retrieved company identifiers were used to manually search for any issued bonds under Datastream's "related securities" link. Selecting that link leads to the list of bonds issued by that company and allows the download of all relevant information such as the name, security identifier code, type of bond, and currency of the bond. Regarding financial information, the end of month total return index series was collected for all the securities[6]. This process was repeated to include firms that are no longer public or non-surviving companies, resulting in a final dataset free of survivorship bias. Finally, ESG scores were collected for the companies in the final dataset.

Following Hoepner and Nilsson (2017), only bonds issued in USD and for the US or international markets were included. All bonds issued in other currencies or markets were excluded from the dataset. Furthermore, a company was excluded if no information was displayed under the "related securities" link, meaning the firm had not issued bonds or they were not covered by Datastream. Bonds with nonstandard characteristics, namely convertible bonds, index-linked bonds, floating-rate notes, and zero-coupon bonds were not excluded from the dataset. Moreover, bond-issuing firms from all sectors were considered. This approach contrasts with most of the previous literature, which excludes financial institutions and banks from their datasets as

[5] The period under analysis is motivated exclusively by the data coverage offered by Thomson Reuters ESG.
[6] Total return data is preferred rather than using bond prices since it takes interest payments and capital gains into account.

companies from the financial sector are typically associated with a substantially larger volume of issued fixed-income securities relative to other industries[7]. Likewise, former studies usually exclude bonds with nonstandard characteristics (Hoepner & Nilsson, 2017; Oikonomou, Brooks, & Pavelin, 2014)[8]. In a real-life scenario, diversified investors' portfolios comprise all types of fixed-income securities, regardless of their characteristics, which is the main motivation behind the decision to include them in the dataset. So, from the 851 firms and 8670 bonds in the sample, 161 firms and 3783 bonds belong to the Financial industry. The Consumer Discretionary and Industrials sectors account for 133 and 131 companies, as well as 765 and 1192 bonds, respectively. The tables reporting the distribution of fixed-income securities within industries and bond types can be consulted in Appendices A and B, respectively.

4.2. Corporate social responsibility scores

Several agencies like Thomson Reuters ESG, Sustainalytics, or KLD, focus on the evaluation of CSR using information from several sources, such as media, company reports, the government, or even third-party providers.

Corporate social responsibility scores are retrieved from the Thomson Reuters ESG Scores database, which is an enhancement and replacement of the previously existing ASSET4 ESG database. This database provides information relative to the assessment of firms' ESG performance worldwide, starting in 2002, and includes scores for three individual pillars (environmental, social, and governance) as well as an ESG combined score. Furthermore, it aims to measure a firm's relative ESG performance, efficiency, and commitment in an objective way, across 10 main subjects such as human rights, environmental product innovation, emissions, and shareholders, as well as 400 ESG metrics. This database offers data regarding ESG ratings on over 9000 companies, globally (Environmental, social and governance (ESG) scores from Refinitiv, 2020, p. 5).

The Thomson Reuters ESG database provides CSR scores based on an aggregate ESG measure, as well as on each of the individual dimensions. Ratings take values between 0 and 100. The ESG combined score measures the overall ESG performance of firms using publicly reported

[7] Previous literature argues that the inclusion of banks and financial institutions could be responsible for a large decrease of cross-industrial variability and possibly skew the results (Hoepner & Nilsson, 2017; Oikonomou et al., 2014). The inclusion of these firms in the sample is motivated by the attempt to replicate an investor's universe of investment possibilities as accurately as possible.

[8] The inclusion of fixed-income securities with nonstandard characteristics is motivated by the attempt to replicate an investor's universe, as well.

data. The environmental pillar measures the performance of a company relative to its innovation, emissions, and resource use. The social dimension assesses a firm's performance relative to human rights, workforce, product responsibility, and community, thus representing a reflection of the company's reputation, as well as its ability to build trust in its workforce. The governance pillar assesses a company's CSR strategy, management performance, and its relationship with shareholders. Additionally, this database provides an ESG controversies score, which aggregates controversies across 10 different categories into a unique category score. Finally, the ESG combined score "overlays the ESG score with ESG controversies to provide a comprehensive evaluation on the company's sustainability impact and conduct" (Thomson Reuters ESG Scores from Refinitiv, 2019, p. 6).

This database has been vastly used in previous SRI performance evaluation studies such as Halbritter and Dorfleitner (2015), Stellner, Klein, and Zwergel (2015), and Gonenc and Scholtens (2017). Using Thomson Reuters ESG carries some advantages relative to other databases, such as its wide coverage, reporting consistency, and transparency of information. Data is collected from numerous sources such as company websites, annual reports, CSR reports, or NGO websites, thus assuring transparency of information. Moreover, Thomson Reuters also provides financial data on the same companies, thus preventing matching errors from collecting information about the companies from different sources of data (Gonenc & Scholtens, 2017). This database faces a limitation concerning rewriting history, as only CSR ratings previous to the latest five fiscal years are marked as definitive (Environmental, social and governance (ESG) scores from Refinitiv, 2020, p. 5). However, there is not enough empirical evidence to assess the impact of this limitation on SRI performance evaluation studies.

4.3. Descriptive statistics

4.3.1. Dataset and portfolio descriptive statistics

The final dataset is composed of 8670 bonds, issued by 851 US companies. Of these firms, 20 are non-surviving and responsible for the issuance of 65 fixed-income securities. Thus, the dataset is free of survivorship bias.

The descriptive statistics of the company CSR scores retrieved from Datastream[9] are reported in Table 1, as follows:

Table 1. Descriptive statistics on ESG scores

	Environmental Score	Social Score	Governance Score	ESG combined Score
Mean	45.192	49.426	51.483	48.614
Standard deviation	18.574	16.931	17.057	14.722
Median	42.025	47.557	52.095	47.66
Minimum	7.8	10.36	9.697	12.79
Maximum	93.719	93.634	92.619	87.782
Skewness	0.399	0.302	-0.144	0.328
Kurtosis	2.23	2.483	2.444	2.379
Observations	851	851	851	851

This table presents the descriptive statistics on the individual and combined ESG scores of all companies included in the dataset over the period from 2003 until 2018. The mean, standard deviation, median, maximum, minimum, skewness, and kurtosis are reported.

The mean CSR rating is similar for all the individual pillars, as well as for the ESG combined score, since all average scores range between 45 and 52. The governance pillar shows the highest average score of about 51, while the average environmental dimension score is the lowest (close to 45). According to Halbritter and Dorfleitner (2015), average CSR ratings are expected to be close to 50, due to the approach followed by Thomson Reuters ESG. Thus, firms in the dataset show corporate social responsibility practices that meet the average. Apart from the corporate governance score, firms in the dataset present moderately right-skewed ratings for the social, environmental, and ESG combined scores. Finally, regarding normality, none of the CSR ratings

[9] The ESG scores were collected on January 31st, 2020.

follow a normal distribution. A table of correlations for the ESG ratings is presented in Appendix C, showing correlation coefficients above 0.700 between the ESG combined score and the individual pillars. Furthermore, the environmental and social ratings are considerably correlated, as well, with a correlation coefficient of 0.702.

Based on the CSR scores of the companies in the dataset, high-rated, low-rated and difference bond portfolios were constructed, following a positive screening investment strategy. The respective descriptive statistics are reported in Table 2.

Table 2. Descriptive statistics on ESG scores for positive screening investment strategy bond portfolios

	Environmental Score		Social Score		Governance Score		ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	92.73	38.81	88.12	42.94	86.73	38.30	84.98	45.59
Standard Deviation	4.52	12.55	4.74	12.04	4.79	9.81	4.47	10.85
Minimum	58.38	5.71	65.93	5.88	72.95	4.49	63.79	11.34
Maximum	99.09	61.37	99.04	64.75	98.79	53.8	97.66	63.64
Median	92.35	37.85	88.1	44.14	86.81	39.57	85.81	46.2
Skewness	-2.378	-0.028	-0.422	-0.245	-0.17	-0.755	-1.142	-0.301
Kurtosis	16.14	2.06	3.69	1.89	2.52	3.07	5.96	2.15
Observations	125886	117459	121887	116861	119304	120092	121106	117126

This table presents the descriptive statistics on the ESG scores from positive screening portfolios, for individual and aggregate scores. The high-rated portfolios are composed by the bonds issued by the 25% top-performing companies, whereas the low-rated portfolios consist bonds issued by firms with the 25% lowest scores. By order, the statistics showed in the table are the mean, standard deviation, minimum, maximum, median, skewness, kurtosis, and the number of observations for each portfolio. The cut-off rate is 25%.

Table 2 presents some summary statistics of the portfolios constructed using a positive screening investment strategy, with a 25% cut-off rate. It is possible to observe that firms included in the high-rated portfolios present, on average, scores greater than 84. Regarding low-rated portfolios, average ratings vary between 38 and 46. However, companies in the low-rated portfolios show average ratings close to the dataset mean ratings, presented in Table 1, particularly for the social and ESG combined scores. Furthermore, all scores are negatively skewed, which means that most of the firms included in the portfolios reveal solid corporate practices. To ensure a larger spread between the ESG scores of the high- and low-rated portfolios, alternative cut-offs of 10% and 5% are also analyzed and summary statistics are reported in Appendices D and E, respectively. As expected, the average CSR ratings generally increase for the high-rated portfolios and decrease for the low-rated portfolios, when social screening intensifies. For the 10% cut-off rate, average ratings for high-rated portfolios range between 88 and 96, whereas for low-rated portfolios mean scores vary between 27 and 36. Regarding the 5% cut-off rate, average ratings for the high- and low-rated portfolios vary between 88 and 97, and 22 and 31, respectively. Descriptive statistics for the returns of the positive screening investment strategy bond portfolios are shown in Table 3. As in Table 2, a cut-off rate of 25% is used.

Table 3. Descriptive statistics on returns for positive screening investment strategy bond portfolios

	Mean	Standard Deviation	Median	Minimum	Maximum	Skewness	Kurtosis	*p*-value
Env. Score								
High-rated	0.450%	2.184%	-10.990%	9.613%	0.546%	-0.637	8.449	0.0000
Low-rated	0.490%	2.058%	-10.389%	7.884%	0.470%	-0.591	7.744	0.0000
Difference	-0.039%	0.986%	-4.242%	3.143%	0.012%	-0.386	6.005	0.0000
Soc. Score								
High-rated	0.443%	2.076%	-8.975%	6.877%	0.441%	-0.432	6.334	0.0000
Low-rated	0.546%	2.140%	-11.539%	7.694%	0.503%	-0.908	9.133	0.0000
Difference	-0.103%	0.925%	-2.551%	4.562%	-0.101%	0.745	7.151	0.0000
Gov. Score								
High-rated	0.418%	1.871%	-7.535%	5.315%	0.458%	-0.612	4.905	0.0000
Low-rated	0.486%	2.011%	-9.854%	7.212%	0.579%	-0.600	7.312	0.0000
Difference	-0.068%	0.738%	-2.859%	3.455%	-0.065%	-0.026	6.995	0.0000
ESG Score								
High-rated	0.372%	1.925%	-7.629%	5.243%	0.427%	-0.719	5.257	0.0000
Low-rated	0.507%	2.098%	-11.236%	7.543%	0.539%	-0.831	8.844	0.0000
Difference	-0.135%	0.972%	-3.742%	3.991%	-0.084%	-0.144	6.108	0.0000

This table presents the descriptive statistics on monthly bond portfolio returns for the positive screening investment strategy, between 2003 and 2018. Portfolios are value-weighted and the chosen cut-off rate is 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. By order, the statistics showed in the table are the mean, standard deviation, median, minimum, maximum, skewness, kurtosis, and the *p*-value for the Jarque-Bera test.

Regarding bond portfolio returns, high and low-rated portfolios based on all CSR scores present positive average returns. However, the average returns of low-rated portfolios are greater than those of high-rated portfolios. These results are consistent for all CSR scores. Furthermore, the return spread between high- and low-rated portfolios is greater for social and ESG combined screens than for portfolios based on environmental and governance ratings. Regarding skewness, returns are negatively skewed (except for the low-rated social portfolio). Finally, the *p*-values resulting from the Jarque-Bera test allow for a strong rejection of the normality hypothesis.

The performance analysis in this dissertation is conducted following not only a positive screening strategy but also a best-in-class strategy. Therefore, descriptive statistics for the returns of high-rated, low-rated, and difference portfolios for this screening strategy are presented in Table 4, for a cut-off rate of 25%.

The bond portfolio mean excess returns reported in Table 4 show positive returns for all high- and low-rated portfolios, regardless of the CSR score. Similar to the findings from Table 3, low-rated portfolios present greater average excess returns than high-rated portfolios. Furthermore, returns for all high- and low-rated portfolios are negatively skewed. Again, the *p*-values from the Jarque-Bera allow a strong rejection of the hypothesis of normality.

Table 4. Descriptive statistics on returns for the best-in-class screening investment strategy bond portfolios

	Mean	Standard Deviation	Median	Minimum	Maximum	Skewness	Kurtosis	*p*-value
Env. Score								
High-rated	0.474%	2.255%	-10.865%	9.281%	0.494%	-0.438	7.411	0.0000
Low-rated	0.515%	1.995%	-10.215%	6.895%	0.486%	-0.748	7.776	0.0000
Difference	-0.041%	0.916%	-3.643%	3.877%	-0.008%	0.455	7.181	0.0000
Soc. Score								
High-rated	0.479%	2.092%	-8.500%	6.011%	0.495%	-0.515	5.769	0.0000
Low-rated	0.523%	2.070%	-11.049%	7.758%	0.463%	-0.846	8.763	0.0000
Difference	-0.044%	0.779%	-2.080%	3.962%	-0.043%	0.877	7.457	0.0000
Gov. Score								
High-rated	0.455%	2.027%	-8.470%	6.155%	0.509%	-0.530	5.886	0.0000
Low-rated	0.476%	2.083%	-10.022%	7.156%	0.518%	-0.606	7.084	0.0000
Difference	-0.021%	0.690%	-2.320%	4.808%	-0.008%	1.575	16.001	0.0000
ESG Score								
High-rated	0.440%	2.042%	-7.820%	6.595%	0.465%	-0.490	5.281	0.0000
Low-rated	0.533%	2.079%	-11.046%	7.716%	0.420%	-0.782	8.631	0.0000
Difference	-0.093%	0.848%	-3.515%	4.553%	-0.071%	0.491	10.176	0.0000

This table presents the descriptive statistics on monthly bond portfolio returns for the best-in-class screening investment strategy, between 2003 and 2018. Portfolios are value-weighted and the chosen cut-off rate is 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies within each industry, whereas the low-rated portfolios comprise bonds issued by the 25% worst performing firms within each industry. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. By order, the statistics showed in the table are the mean, standard deviation, median, minimum, maximum, skewness, kurtosis, and the *p*-value for the Jarque-Bera test.

4.3.2. Factor returns

The benchmark indexes used in this study belong to the ICE BofA Merrill Lynch family[10]. The bond market factor is calculated as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor is computed as the difference between the ICE BofA US High Yield Index and the ICE BofA US Treasury Index[11]. Regarding the option factor, it is calculated as the spread between the ICE BofA US Mortgage-Backed Securities Index and the ICE BofA US Treasury Index. Lastly, the excess returns of the S&P 500 Composite Index are measured by subtracting the risk-free rate, to obtain the equity factor. Total return index series for all indexes are collected using Datastream, except for the risk-free rate, which was retrieved from Professor Kenneth French's data library[12], and discrete returns are calculated for all risk factors. Descriptive statistics on monthly excess returns of the factors are reported in Table 4.

[10] The choice of this family of indices was motivated by the availability of the total return index series.

[11] The default spread factor can also be calculated as the return spread between a BAA-rated corporate bond yield and an AAA-rated corporate bond yield.

[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Consulted on February 20th.

Table 5. Descriptive statistics on monthly excess returns of the the risk factors

	Bond	Default	Option	Stock
Mean	0.310%	0.395%	0.030%	0.694%
Standard Deviation	1.573%	3.020%	0.733%	3.874%
Median	0.391%	0.475%	0.033%	1.044%
Minimum	-7.471%	-16.153%	-2.399%	.16.875%
Maximum	5.595%	13.365%	3.339%	10.929%
Skewness	-0.989	-0.994	0.200	-0.769
Kurtosis	8.840	11.199	5.962	5.249
Jarque-Bera	304.100	569.400	73.310	59.380
***p*-value**	0.000	0.000	0.000	0.000
Observations	192	192	192	192

This table presents the summary statistics on the risk factors, over the period 2003-2018, used in the performance evaluation models. By order, the reported statistics, in percentage, are the mean excess returns, the standard deviation, median, minimum, maximum, skewness, kurtosis. Furthermore, the Jarque-Bera normality test statistic and the respective probability of the test (*p*-value) are presented.

Mean excess returns are similar for all risk factors, between 2003 and 2018. The option factor presents the lowest mean excess returns, whereas the highest mean excess returns are displayed by the default factor. The equity factor displays a substantially larger standard deviation when compared to the bond market factor, which is expected, as most of the risk associated with bonds can be eliminated through diversification (Derwall & Koedijk, 2009). Overall, the risk factors present low correlation coefficients, except for the default spread factor, which is somewhat positively correlated with the option and stock market factors (0.626 and 0.704, respectively)[13]. Finally, the large Jarque-Bera coefficients and the *p*-values equal to zero allow the rejection of the normality hypothesis for all factor return series, meaning none of the factors are normally distributed.

[13] A table of correlations between the risk factors can be consulted in Appendix F.

4.3.3. Public Information variables

A set of three predetermined information variables is used in this research, namely the term spread, the short-term rate, and the inverse relative wealth. The term spread is measured as the spread between the US 10-year Constant Maturity rate and the US Treasury Constant Maturity 3-month yield. The US Treasury Constant Maturity 3-month yield is used as the short-term rate. Finally, the IRW is computed as the exponentially weighted average of the S&P 500 Composite, to proxy for past levels of wealth, deflated by the US Consumer Price Index Urban: All Items[14]. Following Ilmanen (1995), a smoothing coefficient of 0.9 and a 36-month window is used to calculate the IRW. The smoothing parameter is applied to capture business cycle effects. Moreover, the PIV are one-period lagged and data for all variables were collected from Datastream.

The information variables are stochastically detrended by subtracting a 12-month trailing moving-average, as in Silva et al. (2003), Ayadi and Kryzanowski (2011), and Leite and Cortez (2018). The variables are treated in such a manner due to non-stationarity problems usually associated with them, as well as to avoid spurious regression biases. Furthermore, to reduce possible scale effects on the results, the PIV were used in their zero-mean forms. Summary statistics for the PIV in both regular and mean-zero values are reported in table 6.

[14] The IRW and the term spread are used as lagged information variables to proxy for the expected bond risk premium. However, both incorporate information about the expected nominal or real interest rate, as well as the bond risk premium, meaning neither variable is a perfect proxy (Ilmanen, 1995).

Table 6. Descriptive statistics on the public information variables

	Term Spread	Term Spread (Detrended)	Short-term Rate	Short-term Rate (Detrended)	IRW	IRW (Detrended)
Mean	1.904%	0.000%	1.258%	0.000%	0.957	0.000
Median	2.050%	-0.077%	0.400%	-0.006%	0.928	0.00179
Standard Deviation	1.053%	0.524%	1.586%	0.555%	0.114	0.0979
Minimum	-0.600%	-1.006%	0.000%	-2.210%	0.836	-0.306
Maximum	3.790%	1.567%	5.160%	1.018%	1.574	0.384
Skewness	-0.408	0.469	1.292	-1.310	2.863	0.281
Kurtosis	2.480	3.105	3.396	6.422	12.560	6.912
Jarque-Bera	7.489	7.141	54.69	148.6	992.7	125
***p*-value**	0.024	0.028	0.000	0.000	0.000	0.000
Observations	192	192	192	192	192	192

This table reports, in percentage, the summary statistics on the monthly returns of the term spread, the short-term rate, and the IRW. The information variables are 1-month lagged, stochastically detrended using a 12-month trailing moving average and transformed into their mean-zero values. Variables are presented in both original and detrended forms.

The information variables present average positive values for the entire sample period. As expected, in the detrended form, mean excess returns are equal to zero. The standard deviations decreased for all variables after the transformation, especially for the short-term rate. Furthermore, the normality hypothesis is rejected for all information variables, both in the regular and detrended forms. Lastly, the term spread and the short-term rate, in their detrended forms, are somewhat negatively correlated, by -0.733, which is expected as both variables are measured with the US Treasury Constant Maturity 3-month rate[15].

[15] The correlations between the PIV in their detrended forms are presented in Appendix G.

5. Empirical results

This section presents the empirical results obtained from the performance analysis of SRI US corporate bond portfolios from January 2003 to December 2018. In the first subsections, the performance is assessed for the positive approach using conditional and unconditional multi-factor models, followed by the results for the best-in-class screening strategy. Then, additional results for equally-weighted portfolios, alternative cut-offs, and the exclusion of the financial sector are reported, to study whether the portfolio performance remains robust. Finally, financial performance is analyzed over time.

5.1. Positive screening approach

5.1.1. Unconditional model

This section presents and discusses the results obtained from regressing bond portfolio returns against the risk factors. The analysis of the alphas allows for the detection of any abnormal returns. Thus, if alpha is statistically significant, one can assume the presence of abnormal returns. When the alpha is positive and statistically significant, evidence supports that the ESG screened portfolio performs better than the benchmark. Conversely, a negative and statistically significant alpha shows underperformance of the SRI portfolio. Furthermore, if alpha is statistically significant for the difference portfolio, there is evidence of differences between the performance of high- and low-rated portfolios.

The empirical results of the unconditional multi-factor model are summarized in Table 6, for both individual and aggregate ESG scores. The model used in this section is the one given by Equation 4, as described in the methodology section. Portfolios are constructed following a positive screening investment strategy and the chosen cut-off rate is 25%. Moreover, high-rated, low-rated, and difference portfolios are analyzed.

Table 7. Unconditional performance evaluation model regression outputs for individual and ESG combined scores – positive screening investment strategy

	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$
Env. Score						
High-rated	0.0577%	1.1105***	0.0688	-0.4837*	0.0515	0.8317
Low-rated	0.1148%	0.9829***	0.1034	-0.4144*	0.0603*	0.7896
Difference	-0.0570%	0.1276	-0.0345	-0.0693	-0.0088	0.0391
Soc. Score						
High-rated	0.0794%	1.0380***	0.0340	-0.5126**	0.0633*	0.8136
Low-rated	0.1561%**	1.0575***	0.1205**	-0.4941**	0.0420	0.8319
Difference	-0.0767%	-0.0195	-0.0865*	-0.0185	0.0214	0.0412
Gov. Score						
High-rated	0.1156%*	1.0082***	-0.0457	-0.4662**	0.0318	0.8538
Low-rated	0.1007%*	1.0889***	0.0564	-0.2413	0.0473*	0.8775
Difference	0.0149%	-0.0807**	-0.1021***	-0.2248**	-0.0155	0.4530
ESG Score						
High-rated	0.0470%	1.0380***	-0.0276	-0.4808**	0.0408*	0.8701
Low-rated	0.1203%	0.9992***	0.1299**	-0.4986**	0.0583**	0.8209
Difference	-0.0734%	0.0388	-0.1575***	0.0178	-0.0176	0.2667

This table presents the regression estimates of monthly abnormal returns (alphas) for the unconditional multi-factor model, for the portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted coefficient of determination of the regression, R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, designed by *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and value-weighted. The bond portfolios are formed based on a cut-off rate of 25%. The high-rated portfolios are composed of the bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

The regression outputs from the unconditional model show that all high- and low-rated portfolios present positive alphas, although not all statistically significant. Evidence suggests that the low-rated portfolio based on social scores shows positive abnormal returns of 0.1561% (at the 5% level). Concerning environmental and ESG combined screens, no portfolios yield alphas with statistical significance, whereas governance screens present alphas with significance for high- and low-rated portfolios, although only at the 10% level. When compared to previous literature on bond portfolio performance, results regarding portfolios based on environmental scores contrast with Hoepner and Nilsson (2017), who find evidence of significant underperformance relative to the market for high- and low-rated portfolios.

Regarding the risk factors, all factors seem to have some explanatory power over returns, especially the bond and option factors, which are statistically significant for most of the portfolios. The bond market factor shows the highest coefficients, as expected, with statistical significance for the high and low-rated portfolios (at the 1% and 5% levels), regardless of the rating. When it comes to the default factor, however, the coefficients suggest less explanatory power over returns. Concerning the high-rated portfolios, the default factor shows coefficients that are not statistically significant for any of the individual scores. Furthermore, low-rated portfolios load more heavily on this factor, in comparison to high-rated portfolios. Regarding the option factor, evidence suggests that this factor has a negative and significant influence on portfolio returns. Moreover, high-rated portfolios based on individual scores are considerably more exposed to this factor than low-rated portfolios (except for the portfolios formed on the ESG combined score). Finally, the equity factor does not show great explanatory power over returns.

The overall regression outputs from the unconditional performance evaluation model do not show significant evidence supporting that an investment strategy based on CSR ratings yields abnormal returns for investors, except for the low-rated portfolio based on social scores. These results contrast with Polbennikov et al. (2016), who find evidence of abnormal returns for high-rated portfolios. Furthermore, there is no evidence of any differences between the performance of high- and low-rated portfolios.

As previously mentioned in the Methodology section, unconditional performance evaluation models have some limitations, as they do not account for changes in the state of the economy. Since this drawback can lead to incorrect performance estimates, the following section analyzes the performance of SRI bond portfolios using a conditional multi-factor model.

5.1.2. Conditional model

This section presents the results obtained from regressing bond portfolio excess returns against the conditional risk-factors. Conditional models consider the state of the economy, which is proxied by public information variables, and allow for time-varying alphas and betas, thus providing more reliable estimates of financial performance.

The regression outputs for the conditional model, as expressed in Equation 9, are presented in Table 8 for high-rated, low-rated, and difference portfolios. A cut-off rate of 25% is used and portfolios are formed based on aggregate and individual ESG scores. Furthermore, the probability values for the Wald tests of significance for time-varying alphas, time-varying betas, and joint significance of time-varying alphas and betas are presented, as well.

Table 8. Conditional performance evaluation model regression outputs for individual and ESG combined scores – positive screening investment strategy – 25% cut-off rate

	$\propto$	$\propto_{TS}$	$\propto_{STR}$	$\propto_{IRW}$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}
Env.Score								
High-rated	0.1195%*	0.0873%	0.1549%	0.2218%	1.2951***	-0.0392	0.2434	0.0129
Low-rated	0.1074%	-0.2012%	-0.0289%	0.9083%	1.0754***	0.1497**	-0.2276	0.0234
Difference	0.0120%	0.2885%	0.1838%	-0.6865%	0.2198**	-0.1889***	0.4709**	-0.0105
Soc. Score								
High-rated	0.0719%	0.1682%	0.2231%	2.7632%*	1.1603***	0.0336	-0.1777	0.0132
Low-rated	0.1646%**	-0.2281%	-0.1370%	0.5408%	1.0802***	0.1729***	-0.4069**	0.0142
Difference	-0.0927%	0.3963%**	0.3601%**	2.2224%***	0.0801	-0.1393**	0.2292	-0.0009
Gov. Score								
High-rated	0.0985%*	-0.1236%	0.0509%	1.2128%*	1.1927***	-0.0132	-0.2095	0.0006
Low-rated	0.1098%**	-0.1892%	-0.1009%	0.5798%	1.2115***	0.0300	0.1188	0.0252
Difference	-0.0113%	0.0656%	0.1518%	0.6330%	-0.0188	-0.0433	-0.3283**	-0.0246
ESG Score								
High-rated	0.0436%	0.0633%	0.2679%*	1.3126%*	1.1995***	-0.0423	-0.1602	0.0170
Low-rated	0.1270%*	-0.1989%	-0.0863%	0.7047%	1.0444***	0.1646***	-0.3303*	0.0229
Difference	-0.0834%	0.2622%	0.3543%*	0.6079%	0.1551**	-0.2068***	0.1700	-0.0059

	$\beta_{Bond*TS}$	$\beta_{Default*TS}$	$\beta_{Option*TS}$	$\beta_{Equity*TS}$	$\beta_{Bond*STR}$	$\beta_{Default*STR}$	$\beta_{Option*STR}$	$\beta_{Equity*STR}$
Env.Score								
High-rated	-0.6087***	-0.0201	0.3774	0.0868	-0.2611	-0.1419	1.0731**	-0.0668
Low-rated	0.2320	-0.0710	0.9374**	0.0688	0.2097	-0.0541	0.5809	0.0198
Difference	-0.8407***	0.0509	-0.5601	0.0180	-0.4708	-0.0879	0.4922	-0.0867
Soc. Score								
High-rated	-0.2792	-0.2068	0.1811	0.1570**	-0.0038	-0.2719	0.7224	-0.0422
Low-rated	0.3588**	-0.0895	0.8674**	0.0782	0.2229	-0.0394	0.6407	0.0199
Difference	-0.6380***	-0.1173	-0.6863*	0.0788	-0.2267	-0.2325	0.0817	-0.0621
Gov. Score								
High-rated	-0.1731	0.0368	0.0749	0.0593	0.1777	0.0424	0.3275	-0.0539
Low-rated	-0.1041	0.0009	0.4012	0.0772*	-0.0662	0.0074	0.3768	-0.0057
Difference	-0.0689	0.0359	-0.3264	-0.0179	0.2438*	0.0350	-0.0493	-0.0482
ESG Score								
High-rated	-0.3946**	0.0254	-0.0015	0.0563	-0.0397	-0.0370	0.3612	-0.0556
Low-rated	0.3090*	-0.1060	0.8304*	0.0871	0.1798	-0.0797	0.6316	0.0273
Difference	-0.7036***	0.1313	-0.8320**	-0.0308	-0.2195	0.0427	-0.2703	-0.0829

	$\beta_{Bond*IRW}$	$\beta_{Default*IRW}$	$\beta_{Option*IRW}$	$\beta_{Euity*IRW}$	$Adj.R^2$	**W1**	**W2**	**W3**
Env.Score								
High-rated	-0.6626	-0.8587***	2.5616*	0.4725**	0.8719	0.8549	0.0000	0.0000
Low-rated	-1.7676**	-0.4718	0.6288	0.2785	0.8500	0.3352	0.0000	0.0000
Difference	1.1050	-0.3870	1.9328	0.1940	0.2335	0.3819	0.0000	0.0000
Soc. Score								
High-rated	-1.7009**	-1.0912**	0.8234	0.6848***	0.8713	0.0095	0.0000	0.0000
Low-rated	-1.5121**	-0.3330	1.4926	0.1935	0.8756	0.4748	0.0000	0.0000
Difference	-0.1889	-0.7583**	-0.6693	0.4913*	0.2695	0.0353	0.0000	0.0000
Gov. Score								
High-rated	-0.8334*	-0.3542	0.7290	0.0544	0.9019	0.0785	0.0000	0.0000
Low-rated	-1.5270***	-0.1815	-0.3395	0.1456	0.9067	0.3780	0.0000	0.0000
Difference	0.6937	-0.1728	1.0684	-0.0912	0.5180	0.4711	0.0007	0.0011
ESG Score								
High-rated	-0.6412	-0.5275*	0.8803	0.1015	0.9030	0.0472	0.0000	0.0000
Low-rated	-1.7961***	-0.4257	0.8210	0.3591	0.8678	0.4674	0.0000	0.0000
Difference	1.1550**	-0.1018	0.0592	-0.2576	0.4063	0.2838	0.0000	0.0000

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and value-weighted. The bond portfolios are formed using a cut-off rate of 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections

In comparison to the unconditional model, discussed in the previous subsection, the adjusted R^2 increased in all portfolios, regardless of the CSR rating. Hence, the inclusion of information variables into the model has a positive effect on its explanatory power.

Concerning the Wald test, probability values for this test show evidence supporting the existence of time-varying betas, as well as the hypothesis of joint time-varying alphas and betas, since the null hypothesis is strongly rejected for all portfolios. However, there is not much evidence of time-varying alphas (apart from the high-rated and difference portfolios based on social ratings and the high-rated ESG combined portfolio). Thus, the results from the Wald test sustain the use of a conditional model to assess corporate bond portfolio performance.

As can be observed in Table 8, the low-rated portfolios based on social and governance screens yield abnormal alphas (at the 5% level). Regarding high-rated portfolios, results show significant abnormal returns when environmental and governance screens are applied, although only at the 10% level. Furthermore, the difference between high- and low-rated portfolios is not statistically significant, similar to the findings from Hoepner and Nilsson (2017), as they conclude that the difference portfolios do not yield abnormal returns for individual screens, for aggregate scores. However, these results are in line with Pereira et al. (2019), who find no differences between the performance of high- and low-rated portfolios. Finally, findings contrast with Polbennikov et al. (2016), who find evidence of abnormal returns for high-rated portfolios.

The alphas for the lagged information variables indicate that the short-term rate, the term spread, and the IRW do not have much explanatory power over portfolio returns, as these variables are not statistically different from zero (except for the difference portfolio based on social ratings).

Regarding factor loadings, results are somewhat similar to the ones obtained from the unconditional multi-factor model (Table 7). Findings suggest that portfolios are highly exposed to the bond factor (mostly at the 1% level), as expected, regardless of the ESG rating. Furthermore, high-rated portfolios are considerably more exposed to the bond factor than low-rated portfolios. The default factor has a positive and significant effect on low-rated portfolios (at the 1% and 5% levels), whereas it negatively and significantly influences long-short portfolios (at the 1% and 5% levels, as well). However, this risk factor does not show any significant results for portfolios based on the governance score. The exposure to the option factor varies across portfolios. While for portfolios based on environmental screens, only the difference portfolio is positively exposed to this factor (at the 5% level), all other statistically significant coefficients negatively affect portfolio returns (at the 10% or 5% levels). Finally, some changes can be documented relative to the unconditional model exposure to the equity factor, as it presents relatively low coefficients, which are not statistically different from zero, for all portfolios.

Overall, findings from Table 8 show positive and statistically significant abnormal returns for the low-rated portfolios based on social and governance ratings, at the 5% level. However, none of the difference portfolios present statistically significant alphas. Thus, there is no evidence in the results supporting that high-rated portfolios perform differently from low-rated bond portfolios.

As mentioned before, this study aims to evaluate the financial performance of corporate bond portfolios, using both the positive and best-in-class approaches. So, the following section presents the results for the conditional model regression outputs when a best-in-class investment strategy is applied to construct the bond portfolios.

5.2. Best-in-class screening investment approach

The best-in-class screening approach is applied to assess whether results remain robust to industry differences. Following this approach, portfolios are constructed by including the bonds issued by the best and worst-performing firms within each industry. Using a 25% cut-off rate, high-rated, low-rated, and difference portfolios are constructed, based on individual and aggregate ESG ratings.

Table 9 shows the conditional regression outputs estimated for the best-in-class screening investment strategy. As can be observed, the adjusted R^2 for high and low-rated portfolios ranges between 87% and 92%. Regarding the Wald tests, results support the existence of time-varying alphas only for high-rated (excluding the portfolio formed based on environmental screens) and difference portfolios. Furthermore, evidence of time-varying betas as well as joint significance of time-varying alphas and betas is present across all portfolios, regardless of the ESG rating.

In terms of conditional alphas, results show that all low-rated portfolios yield positive and statistically significant abnormal returns (at the 5% and 1% levels). Conversely, there is no evidence significant abnormal returns on high-rated portfolios. Moreover, the alphas increased for all low-rated portfolios, relative to Table 8.

As expected, all portfolios heavily load on the bond factor (at the 1% level) and, once again, high-rated portfolios are more exposed to this factor than low-rated portfolios (apart from portfolios based on governance screens). Concerning the default factor, most of the betas are positive, yet not statistically significant. The option factor has a significant effect in only one portfolio (although only at the 10% level) and the coefficients are mostly negative. Finally, none of the portfolios are significantly influenced by the equity factor, and exposure to this factor is low across all portfolios.

Hence, findings suggest that low-rated portfolios seem to provide statistically significant abnormal returns for investors. Furthermore, the regression outputs show a negative and statistically significant alpha for the difference portfolio based on the ESG combined score (at the 5% level) and to a less extent for the one based on the social score (at the 10% level), suggesting that the low-rated portfolios based on these screens perform significantly better than the high-rated portfolios. This empirical evidence contrasts with Kempf and Osthoff (2007), as their results show that the highest alphas are obtained from the best-in-class screening strategy when the portfolios are composed of stocks from top-performing firms. Overall, results are not robust to industry differences, since there is evidence of differences between the performance of bonds issued by top-performing firms and bonds issued by firms with lower ESG ratings, for portfolios based on social and ESG combined ratings.

To assess the robustness of results, the following section presents the regression outputs for alternative cut-offs and weighting schemes, as well as the conditional regression outputs when the financial sector is excluded.

Table 9. Conditional performance evaluation model regression outputs for individual and ESG combined scores – best-in-class approach – 25% cut-off rate

	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W1	W2	W3
Env. Score									
High-rated	0.0858%	1.2988***	0.0149	0.0452	0.0055	0.8902	0.1369	0.0000	0.0000
Low-rated	0.1757%***	1.1064***	0.0733	-0.1230	0.0169	0.8701	0.1919	0.0000	0.0000
Difference	-0.0899%	0.1924***	-0.0583	0.1682	-0.0114	0.3155	0.0369	0.0000	0.0000
Soc. Score									
High-rated	0.0796%	1.2476***	0.0175	-0.0852	0.0160	0.9017	0.0057	0.0000	0.0000
Low-rated	0.1718%***	1.0928***	0.1010*	-0.2912*	0.0284	0.8870	0.5417	0.0000	0.0000
Difference	-0.0921%*	0.1548***	-0.0835*	0.2060	-0.0124	0.3358	0.0079	0.0000	0.0000
Gov. Score									
High-rated	0.0732%	1.2040***	0.0564	-0.1615	0.0064	0.8925	0.0098	0.0000	0.0000
Low-rated	0.1197%**	1.2191***	0.0218	-0.0110	0.0223	0.9111	0.1436	0.0000	0.0000
Difference	-0.0465%	-0.0150	0.0345	-0.1505	-0.0160	0.2819	0.0040	0.0000	0.0000
ESG Score									
High-rated	0.0575%	1.2196***	0.0219	-0.1553	0.0193	0.8864	0.0048	0.0000	0.0000
Low-rated	0.1793%***	1.1036***	0.0768	-0.2704	0.0227	0.8853	0.2509	0.0000	0.0000
Difference	-0.1218%**	0.1159*	-0.0550	0.1151	-0.0034	0.3619	0.0020	0.0000	0.0000

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced, value-weighted, and constructed following a best-in-class screening strategy. The bond portfolios are formed using a cut-off rate of 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

5.3. Robustness tests

This section presents the results for the robustness tests applied to assess if there is any variation in results when different portfolio construction processes are used. The first subsection shows the findings for bond portfolios formed using different cut-offs (10% and 5%). The second subsection reports the results for the conditional model using a different portfolio weighting scheme (equally-weighted portfolios). Finally, the portfolio excess returns are regressed against the risk factors, once again, without including the financial sector, considering a cut-off rate of 25%.

5.3.1. Alternative cut-offs

Thus far, all results for the positive screening investment approach were obtained using portfolios constructed with a 25% cut-off rate. This section focuses on the analysis of portfolios consisting of bonds within the top and bottom 10% and 5% ESG ratings of bond-issuing companies, for each year, to assess whether findings depend on the bond portfolio construction process. This robustness test is conducted for both positive and best-in-class screening strategies.

Some previous literature shows significant abnormal returns only for extreme cut-off rates. For example, Auer (2016) finds that portfolios based on environmental and social scores show positive and statistically significant abnormal returns only when cut-off rates of 5% are used. Conversely, the results obtained by Halbritter and Dorfleitner (2015) are robust to alternative cut-offs. Regression outputs for the conditional model using value-weighted portfolios using 10% (Panel A) and 5% (Panel B) cut-off rates are presented in Table 10.

Concerning the R^2, Table 10 shows a general decrease relative to findings from the 25% cut-off, for both panels. However, for the 10% cut-off, the R^2 for both high- and low-rated portfolios remain above 80% (except for the low-rated portfolio based on the environmental rating).

Findings from the Wald tests are similar to the ones observed in Table 8, as the probability values suggest there is no evidence of time-varying alphas (except for the high-rated portfolios based on the governance score). Furthermore, results for this test still support the presence of time-varying betas, as well as the joint significance of time-varying alphas and betas.

Regarding the coefficients of the risk factors, again, the bond factor has a positive and statistically significant effect on portfolio returns, for both cut-off rates. In terms of exposure, high-rated portfolios are more influenced than low-rated portfolios. Exposure to the default factor is similar to previous findings, whereas the option factor coefficients are mostly insignificant (for Panel B). Finally, the equity factor does not show any statistically significant betas for most portfolios.

Results from both panels of Table 10 show positive alphas for all high- and low-rated portfolios, although in most cases the alphas are not statistically significant. Furthermore, alphas of high-rated portfolios generally increased for both cut-off rates, relative to Table 8, whereas alphas of low-rated portfolios decreased. Overall, as the cut-off rates become more extreme, more evidence of abnormal returns for high-rated portfolios appears, while evidence of abnormal returns for low-rated portfolios fades. In fact, for the 5% cut-off rate, only portfolios composed of bonds issued by top-performing firms present positive and significant alphas (except for the environmental screen). However, global conclusions from this robustness test do not considerably change relative to results from the 25% cut-off, since returns for the difference portfolios are not statistically different from zero. Thus, there is no evidence of differences between the performance of bonds issued by high- and low-rated firms. These findings are in line with Halbritter and Dorfleitner (2015), and Pereira et al. (2019), in terms of long-short portfolios.

Table 10. Conditional performance evaluation model regression outputs for individual and ESG combined scores – positive screening investment strategy – 10% and 5% cut-off rates

Panel A	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	**W1**	**W2**	**W3**
Env. Score									
High-rated	0.0875%	1.3044***	-0.0915*	0.1565	0.0428	0.8446	0.6824	0.0000	0.0000
Low-rated	0.1179%	0.9509***	0.1598*	-0.2179	0.0445	0.7685	0.6066	0.0002	0.0000
Difference	-0.0304%	0.3535***	-0.2514***	0.3744	-0.0017	0.3923	0.2740	0.0000	0.0000
Soc. Score									
High-rated	0.1089%*	1.1878***	-0.1029**	-0.1846	0.0224	0.8582	0.1360	0.0000	0.0000
Low-rated	0.1466%*	0.9944***	0.3074***	-0.5964***	0.0177	0.8336	0.7298	0.0000	0.0000
Difference	-0.0377%	0.1934**	-0.4103***	0.4117*	0.0047	0.5205	0.4489	0.0000	0.0000
Gov. Score									
High-rated	0.1253%*	1.2663***	-0.0704	0.0369	0.0122	0.8625	0.0101	0.0000	0.0000
Low-rated	0.0754%	1.0707***	0.0751	-0.1809	0.0464*	0.8390	0.2165	0.0000	0.0000
Difference	0.0499%	0.1956**	-0.1454**	0.2178	-0.0343	0.2474	0.7283	0.0001	0.0001
ESG Score									
High-rated	0.1228%*	1.1594***	-0.0786	-0.1849	0.0293	0.8451	0.2644	0.0000	0.0000
Low-rated	0.1199%	1.0024***	0.2447***	-0.4613**	0.0525	0.8158	0.6673	0.0000	0.0000
Difference	0.0029%	0.1570	-0.3232***	0.2764	-0.0232	0.4021	0.8308	0.0000	0.0000

Table 10. (Continued)

Panel B	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W1	W2	W3
Env. Score									
High-rated	0.1263%	1.2938***	-0.1534**	0.1667	0.0424	0.8012	0.7320	0.0000	0.0000
Low-rated	0.1064%	0.8898***	0.1074	-0.1261	0.0387	0.6506	0.5306	0.0071	0.0002
Difference	0.0199%	0.4040***	-0.2609**	0.2928	0.0037	0.2558	0.3227	0.0014	0.0003
Soc. Score									
High-rated	0.1124%*	1.1422***	-0.0834	-0.2466	0.0177	0.8427	0.1824	0.0000	0.0000
Low-rated	0.1246%	0.9664***	0.3943***	-0.6152	0.0393	0.7208	0.9080	0.0002	0.0001
Difference	-0.0122%	0.1759	-0.4777***	0.3685	-0.0216	0.3631	0.8221	0.0001	0.0001
Gov. Score									
High-rated	0.1686%***	1.1689***	-0.1006**	-0.0283	0.0077	0.8464	0.0003	0.0000	0.0000
Low-rated	0.0864%	0.9808***	0.0933	-0.2366	0.0431	0.8114	0.0476	0.0000	0.0000
Difference	0.0822%	0.1881**	-0.1939***	0.2082	-0.0353	0.4427	0.0079	0.0000	0.0000
ESG Score									
High-rated	0.1202%*	1.2438***	-0.1294**	0.0110	0.0307	0.8191	0.2695	0.0000	0.0000
Low-rated	0.0538%	0.9323***	0.2569**	-0.3649	0.0675	0.7407	0.5476	0.0000	0.0000
Difference	0.0664%	0.3115**	-0.3864***	0.3759	-0.0368	0.4295	0.7183	0.0000	0.0000

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and value-weighted. The bond portfolios are formed using cut-off rates of 10% (Panel A) and 5% (Panel B). The high-rated portfolios are composed of bonds issued by the 10% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 10% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

Furthermore, robustness tests were also conducted for the best-in-class screening strategy. Table 11 shows the conditional regression outputs for the 10% (Panel A) and 5% (Panel B) cut-off rates. Similar to previous analyses, high-rated, low-rated, and difference portfolios are formed for both aggregate and individual CSR scores.

As can be observed, the adjusted R^2 decreases for all high- and low-rated portfolios, as the cut-off rate becomes more extreme. However, this decline is more pronounced in low-rated portfolios. Probability values from the Wald test support the presence of time-varying betas, as well as the joint significance of time-varying alphas and betas. Yet, there is no evidence of time-varying alphas, conversely to findings from Table 9.

Regarding exposure to the risk factors, findings for the bond factor are quite similar to the previously discussed ones. Conversely, exposure to the default factor increased for most of the portfolios, relative to Table 9. Furthermore, the option factor presents coefficients which are mostly not significant, and the equity factor does not significantly affect returns. Results are similar for both cut-off rates.

In terms of alphas, results provide evidence of positive and statistically significant abnormal returns (at the 5% and 10% levels) for low-rated portfolios, regardless of the ESG rating or cut-off rate. Conversely, for Panel A, the high-rated portfolios show alphas which, are not statistically different from zero (except for the portfolio based on governance screens). Results from Panel B are, however, somewhat distinct, since there is evidence of abnormal returns for high-rated portfolios (at the 5% and 10% levels, except for environmental screens). Furthermore, alphas of low-rated portfolios are larger than those of high-rated portfolios. In general, findings from this approach are not in line with the ones obtained for 25% cut-off rate (Table 9), since all difference portfolios are statistically not significant, meaning there is no longer evidence of outperformance of low-rated portfolios.

As a further robustness test, financial performance is assessed using different portfolios weighting schemes. So, the following subsection presents results for the conditional regression outputs of equally-weighted portfolios.

Table 11. Conditional performance evaluation model regression outputs for individual and ESG combined scores – best-in-class approach – 10% cut-off rate

	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W1	W2	W3
Env.Score									
High-rated	0.0949%	1.3129***	-0.1005**	0.0963	0.0228	0.8851	0.4448	0.0000	0.0000
Low-rated	0.1588%**	0.9290***	0.1558**	-0.4072*	0.0313	0.7665	0.9426	0.0000	0.0000
Difference	-0.0639%	0.3839***	-0.2562***	0.5036**	-0.0086	0.2380	0.6684	0.0007	0.0005
Soc. Score									
High-rated	0.0990%	1.1907***	0.0455	-0.2811	0.0185	0.8374	0.0084	0.0000	0.0000
Low-rated	0.1847%**	0.9568***	0.3336***	-0.5499**	-0.0042	0.7981	0.9405	0.0000	0.0000
Difference	-0.0858%	0.2339**	-0.2880***	0.2688	0.0227	0.3163	0.1174	0.0000	0.0000
Gov. Score									
High-rated	0.1307%**	1.1527***	-0.0408	-0.1348	0.0197	0.8691	0.1279	0.0000	0.0000
Low-rated	0.1564%**	1.1612***	0.0120	-0.1550	0.0431	0.8457	0.0387	0.0000	0.0000
Difference	-0.0257%	-0.0085	-0.0528	0.0202	-0.0234	0.1259	0.3587	0.0061	0.0132
ESG Score									
High-rated	0.0969%	1.2442***	-0.0869	-0.0474	0.0128	0.8698	0.5880	0.0000	0.0000
Low-rated	0.1757%**	0.9665***	0.2085***	-0.4799*	0.0368	0.7992	0.6668	0.0000	0.0000
Difference	-0.0788%	0.2777**	-0.2954***	0.4325	-0.0240	0.3547	0.4333	0.0000	0.0000

Table 11. (Continued)

Panel B	∝	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W1	W2	W3
Env.Score									
High-rated	0,1146%	1.3037***	-0.0994*	0.0782	0.0211	0.8585	0.3421	0.0004	0.0005
Low-rated	0,2028%**	0.8872***	0.1571**	-0.4223*	0.0217	0.7232	0.9504	0.0528	0.0107
Difference	-0,0883%	0.4165***	-0.2565***	0.5005*	-0.0005	0.2027	0.4369	0.0191	0.0084
Soc. Score									
High-rated	0,1243%**	1.1885***	-0.0794	-0.2064	0.0201	0.8552	0.2660	0.0000	0.0000
Low-rated	0,1657%*	0.9100***	0.3891***	-0.6741**	0.0327	0.7453	0.8391	0.0000	0.0000
Difference	-0,0414%	0.2785**	-0.4685***	0.4678	-0.0127	0.3761	0.9086	0.0030	0.0012
Gov. Score									
High-rated	0,1220%**	1.2136***	-0.0516	-0.0341	0.0164	0.8662	0.0149	0.0000	0.0000
Low-rated	0,1633%**	1.0473***	0.0497	-0.2136	0.0509*	0.7851	0.0916	0.0000	0.0000
Difference	-0,0413%	0.1663**	-0.1013	0.1795	-0.0345	0.1412	0.7250	0.1109	0.2124
ESG Score									
High-rated	0,1152%*	1.2261***	-0.0866	-0.0872	0.0136	0.8589	0.6065	0.0000	0.0000
Low-rated	0,1646%*	0.8972***	0.1760**	-0.6718**	0.0668*	0.7639	0.6854	0.0000	0.0000
Difference	-0,0494%	0.3289***	-0.2626***	0.5846**	-0.0532	0.3131	0.4173	0.0000	0.0000

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced, value-weighted, and constructed following a best-in-class screening strategy. The bond portfolios are formed using cut-off rates of 10% (Panel A) and 5% (Panel B). The high-rated portfolios are composed of bonds issued by the 10% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 10% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

5.3.2. Performance of equally-weighted portfolios

So far, all SRI bond portfolios used in the analysis were constructed by calculating value-weighted returns. However, some authors find that firm size impacts value- and equally-weighted portfolio returns. Thus, to assess if findings are robust to alternative portfolio weighting schemes, this section focuses on the analysis of the conditional model regression outputs of equally-weighted portfolios, with a cut-off rate of 25%.

Table 12 presents the results for the high-rated, low-rated, and difference bond portfolios, formed considering individual and aggregate ESG scores. Regarding the adjusted R^2 and results from the Wald test, the use of equally-weighted portfolios delivers similar results to the value-weighted portfolios, for the 25% cut-off rate. There is no evidence of time-varying alphas (except for the high-rated and difference social portfolios), however, the *p*-values suggest the presence of time-varying betas, as well as the joint significance of time-varying alphas and betas.

In terms of risk factors, in contrast with previous results, low-rated portfolios are more exposed to the bond factor than high-rated portfolios. Furthermore, betas for this factor are significant for all portfolios (mostly at the 1% level). Once again, the equity factor provides betas with no statistical significance.

Generally, alphas increased in comparison to the results from value-weighted portfolios. Moreover, high- and low-rated portfolios yield positive and significant abnormal returns (at the 5% and 1% levels). The improvement in financial performance indicates that portfolios constructed using equally-weighted returns perform slightly better than value-weighted portfolios. Moreover, these findings suggest that bonds issued by smaller firms tend to present greater abnormal returns than bonds issued by larger firms. Similar results were reported by Pereira et al. (2019), regarding equally-weighted portfolios. Furthermore, although the alphas from low-rated portfolios seem to be superior to those from high-rated portfolios (except for governance screens), excess returns from the long-short portfolio are, once again, not statistically significant, in line with the results from value-weighted portfolios.

Table 12. Conditional performance evaluation model regression outputs for individual and ESG combined scores – equally-weighted returns – 25% cut-off rate

	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W1	W2	W3
Env. Score									
High-rated	0.1402%**	1.0026***	-0.0760	0.0098	-0.0137	0.8671	0.3585	0.0000	0.0000
Low-rated	0.1691%**	1.0011***	0.1127	-0.2221	0.0209	0.8106	0.3060	0.0000	0.0000
Difference	-0.0289%	0.0015	-0.1887***	0.2318	-0.0345	0.2204	0.4251	0.0061	0.0056
Soc. Score									
High-rated	0.1373%**	0.9209***	-0.0257	-0.3782**	0.0040	0.8383	0.0184	0.0000	0.0000
Low-rated	0.2222%***	1.0336***	0.1421**	-0.4000**	0.0079	0.8446	0.2384	0.0000	0.0000
Difference	-0.0849%	-0.1126*	-0.1679***	0.0217	-0.0039	0.3394	0.0036	0.0004	0.0000
Gov. Score									
High-rated	0.1623%***	0.9540***	-0.0530	-0.3229**	-0.0096	0.8788	0.2938	0.0000	0.0000
Low-rated	0.1390%***	1.0600***	-0.0055	0.0121	-0.0082	0.8933	0.2923	0.0000	0.0000
Difference	0.0233%	-0.1060**	-0.0475	-0.3350**	-0.0014	0.5879	0.8096	0.0000	0.0000
ESG Score									
High-rated	0.1070%**	0.9163***	-0.0828*	-0.3341***	-0.0034	0.8848	0.1761	0.0000	0.0000
Low-rated	0.1836%***	0.9928***	0.1335**	-0.3177	0.0193	0.8391	0.4093	0.0000	0.0000
Difference	-0.0766%	-0.0765	-0.2163***	-0.0164	-0.0227	0.5355	0.3616	0.0029	0.0008

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and equally-weighted. The bond portfolios are formed using a cut-off rate of 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

5.3.3. Exclusion of the financial sector

To the best of my knowledge, most of the existing literature regarding the financial performance of stocks and bonds excludes the financial sector. Some scholars support their decision of doing so under the argument that including financial institutions and banks could reduce cross-industrial variability and, thus, skew results (Hoepner & Nilsson, 2017; Oikonomou et al., 2014). To assess the impact on results of including the financial sector, a robustness test is conducted by reconstructing those portfolios without considering bonds issued by firms belonging to the financial industry. A cut-off rate of 25% is used and returns are regressed against the conditional risk factors using equation 9.

As can be observed in Table 13, the R^2 of the regressions for all high- and low-rated portfolios range between 83% and 89%, indicating that the explanatory power of returns did not change significantly, comparing to Table 8. Probability values from the Wald test show evidence supporting the existence of time-varying alphas for high-rated portfolios (except for the governance screen). There is evidence of time-varying betas and joint significance of time-varying alphas and betas across all portfolios, as in previous analyses.

Regarding the risk factor loadings, overall results are comparable to the ones from Table 8, as well. The bond factor shows the largest positive influence on portfolios returns (at the 5% and 1% levels) and, once again, high-rated portfolios are more exposed than low-rated portfolios. Conversely, none of the betas for the equity factor are statistically significant. Factor loadings for the option and default factors are also quite similar to the results from the previous sections.

It is relevant to point out that alphas increased (except for the low-rated social portfolio), comparing to Table 8. Furthermore, there is evidence of positive and statistically significant abnormal returns for both high- and low-rated portfolios (at the 5% and 1% levels), unlike regression outputs for the inclusion of the financial sector, which provide more evidence of abnormal returns for low-rated portfolios than high-rated. These results suggest that the inclusion of bonds issued by banks and financial institutions may have penalized portfolio returns, especially for high-rated portfolios, since the alphas of these portfolios increased and gained statistical significance with this robustness test. Although findings provide evidence of abnormal returns for high- and low-rated portfolios, the difference is, once again, not statistically significant.

Table 13. Conditional performance evaluation model regression outputs for individual and ESG combined scores – excluding the financial sector – 25% cut-off rate

	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj. R^2$	W1	W2	W3
Env.Score									
High-rated	0.1574%**	1.2170***	-0.0372	-0.1741	-0.0185	0.8677	0.0270	0.0000	0.0000
Low-rated	0.1275%*	1.0751***	0.1372**	-0.1815	0.0157	0.8340	0.1203	0.0000	0.0000
Difference	0.0299%	0.1419	-0.1744**	0.0074	-0.0343	0.3105	0.8907	0.0000	0.0000
Soc. Score									
High-rated	0.1347%**	1.1771***	0.0022	-0.2227	0.0053	0.8472	0.0002	0.0000	0.0000
Low-rated	0.1587%**	1.0749***	0.1778***	-0.4205***	-0.0091	0.8707	0.1057	0.0000	0.0000
Difference	-0.0241%	0.1022	-0.1756***	0.1978	0.0144	0.3562	0.0843	0.0000	0.0000
Gov. Score									
High-rated	0.1390%**	1.2068***	-0.0306	-0.1574	-0.0092	0.8704	0.0765	0.0000	0.0000
Low-rated	0.1442%**	1.1575***	0.0907	-0.1889	-0.0061	0.8820	0.0078	0.0000	0.0000
Difference	-0.0052%	0.0493	-0.1213**	0.0316	-0.0030	0.4099	0.6582	0.0000	0.0000
ESG Score									
High-rated	0.1284%**	1.1855***	-0.0801	-0.2555	0.0057	0.8692	0.0388	0.0000	0.0000
Low-rated	0.1490%**	1.0515***	0.1458**	-0.2745*	0.0150	0.8558	0.2011	0.0000	0.0000
Difference	-0.0206%	0.1340*	-0.2259***	0.0190	-0.0093	0.4637	0.8617	0.0001	0.0001

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and value-weighted. The bond portfolios are formed using a cut-off rate of 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

5.4. Bond portfolio performance over time

Previous literature has dedicated its attention to the changes in the financial performance of SRI over time. Derwall et al. (2011), Bebchuck et al. (2013), and Pereira et al. (2019), for example, find evidence suggesting that investors should no longer expect abnormal returns when engaging in SRI strategies, as investors should have increased their knowledge about the impact of CSR practices. Thus, SRI financial securities should no longer be mispriced. To assess how the financial performance of the bond portfolios changed over time, we analyze three subperiods, namely 2003-2007, 2008-2012, and 2013-2018. The results are presented in Table 14. Similar to the previous analyses, high-rated, low-rated, and difference portfolios are constructed, using a 25% cut-off rate, and the conditional model from Equation 9 is applied.

First, it is relevant to point out that the adjusted R^2 from the regression outputs for the different subperiods are very high, with most of them lying between 90% and 97%. Regarding results from the Wald tests, there is no evidence of time-varying alphas, time-varying betas, or joint significance of time-varying alphas and betas, during the first subperiod. For the second and third subperiods, conversely, findings are in line with previously discussed results. The regression outputs for all subperiods are presented in Appendix H.

Results for the first subperiod, which spans between 2003 and 2007, show no evidence of abnormal returns for high- or low-rated bond portfolios, as all alphas are statistically equal to zero. Moreover, high-rated portfolios display larger alphas than low-rated portfolios (except for social screens).

Concerning the 2008-2012 subperiod, overall, alphas largely increased in comparison to the 2003-2007 period. Furthermore, alphas of low-rated portfolios are greater than alphas of high-rated portfolios (except for portfolios based on governance ratings), in contrast to results from the first subperiod. In terms of financial performance, findings are considerably different. Both high- and low-rated portfolios provide positive and statistically significant alphas (except for the high-rated portfolio based on environmental screens). Additionally, evidence suggests that low-rated portfolios perform significantly better than high-rated portfolios (at the 1% level) for portfolios based on the ESG combined rating and to a less extent (at the 10% level) for portfolios based on environmental scores. For social and governance screens, there are no differences between the financial performance of high- and low-rated portfolios.

Finally, from 2013 until 2018, the alphas decreased and, for most portfolios, lost their statistical significance. From all the portfolios, only the high-rated portfolio based on environmental ratings and the low-rated portfolios based on social and ESG combined ratings provide positive and statistically significant abnormal returns (at the 5% level). However, the alphas of the difference portfolios are no longer statistically significant.

The results from this section can be confronted with the errors-in-expectations and shunned-stock hypotheses presented by Derwall et al. (2011). According to the errors-in-expectations hypothesis, if financial markets fail to incorporate value-relevant information regarding CSR, SRI stocks may be mispriced. Consequently, the possibility of yielding abnormal returns by investing in SRI stocks arises. However, the authors claim that this outperformance should disappear over time and financial performance should be similar to the market as investors increase their knowledge about SRI. Findings from this study contrast with the errors-in-expectations hypothesis, since there is evidence of abnormal returns only after the first subperiod. Furthermore, the shunned-stock hypothesis argues that values-driven investors are willing to sacrifice financial performance due to the nature of some social values. In other words, these investors shun controversial stocks, creating a shortage in their demand and, therefore, it is expected that stocks from low-rated firms will generate higher returns. This effect should remain prevalent over time, as long as there exists a considerable number of values-driven investors. Again, results are not in line with this hypothesis, since between 2003 and 2007, there is no evidence of abnormal returns for low-rated portfolios. Moreover, while low-rated portfolios yield statistically significant alphas for the 2008-2012 period, and environmental and ESG combined screens even outperform high-rated portfolios, this outperformance fades in the 2013-2018 period. It is also worth mentioning that the second subperiod comprises part of the 2007-2008 financial crisis and its subsequent years. Therefore, the results obtained for the 2008-2012 and 2013-2018 periods may be linked to the financial crisis and its consequences in the markets[16].

Overall, results show that the different approaches applied in this research provide some evidence of abnormal returns for both high-rated and low-rated portfolios. However, almost all difference portfolios present alphas with no statistical significance. This means that although there is no consistent evidence of outperformance for SRI trading strategies, there is also no evidence of

[16] The fact that the first subperiod spans from 2003 and 2007, thus comprising the first year of the US financial crisis, may have possibly affected results as well.

underperformance. In sum, although alphas of low-rated portfolios tend to be greater than those of high-rated portfolios throughout the different approaches, there seem to be no differences between the performance of high- and low-rated portfolios, except for the best-in-class screening strategy. For this approach, there is evidence of outperformance of low-rated portfolios based on social and ESG combined ratings, relative to high-rated portfolios. However, findings are not robust to alternative cut-offs. Likewise, when portfolio performance is assessed over time, there is evidence in favor of the superior performance of low-rated portfolios based on environmental and ESG combined screens, comparing to high-rated portfolios, for the 2008-2012 period. Nevertheless, this outperformance is not consistent over time.

Table 14. Conditional performance evaluation model regression outputs for individual and ESG combined scores – performance over time

	$\propto_{2003-2007}$	$\propto_{2008-2012}$	$\propto_{2013-2018}$
Env.Score			
High-rated	0.1164%	0.3694%	0.0947%**
Low-rated	0.0208%	0.7296%***	0.1021%
Difference	0.0956%	-0.3603%*	-0.0074%
Soc. Score			
High-rated	0.0555%	0.6626%**	0.0732%
Low-rated	0.0586%	0.6807%***	0.1547%**
Difference	0.0031%	-0.0182%	-0.0816%
Gov. Score			
High-rated	0,0950%	0.4283%**	0.0441%
Low-rated	0,0537%	0.4096%**	0.0901%*
Difference	0,0413%	0.0186%	-0.0460%
ESG Score			
High-rated	0.0873%	0.3104%*	0.0604%
Low-rated	0.0171%	0.6875%***	0.1256%**
Difference	0.0702%	-0.3771%***	-0.0652%

This table presents the regression estimates of monthly abnormal returns (alphas) for the conditional multi-factor model, for portfolios formed using both individual and aggregate ESG scores, the risk-factor coefficients (betas), and the adjusted R^2. *Bond* is the bond market factor, which is computed as the excess returns of the ICE BofA US Corporate Index over the risk-free rate. The default spread factor, *Default*, is calculated as the return spread between the ICE BofA US High Yield Index and ICE BofA US Treasury Index. *Option* is computed as the difference between the ICE BofA US Mortgage-backed Securities and the ICE BofA US Treasury Index. The *Equity* variable corresponds to the excess returns of the S&P 500 Composite Index. The risk-free rate used is the one from Professor Kenneth French's database. Portfolios are annually rebalanced and value-weighted. The bond portfolios are formed using a cut-off rate of 25%. The high-rated portfolios are composed of bonds issued by the 25% top-performing companies, whereas the low-rated portfolios comprise the bonds issued by the 25% worst performing firms. The difference portfolio corresponds to the long-short portfolio, which is an investment strategy of going long in the high-rated portfolio and going short in the low-rated portfolio. The period of observation spans from 2003 until 2018 and is divided into three subperiods, namely 2003-2007, 2008-2012, and 2013-2018. *W1*, *W2*, and *W3* represent the *p*-values from the Wald test for time-varying alphas, time-varying betas and the joint significance. Statistical significance is indicated by ***, **, and * at the 1%, 5%, and 10% levels, respectively. Standard errors are corrected for autocorrelation and heteroskedasticity problems when necessary using the Newey-West (1987) and White (1980) corrections.

6. Conclusion

All over the world, interest in SRI has been growing among investors. The possibility of incorporating socially responsible criteria into the investment decision without hurting financial gains is attractive to companies wishing to attain to their social responsibilities. To assess whether such a reality can be accomplished, several studies were conducted. However, most literature focuses on the performance evaluation of equity securities, leaving the fixed-income area less explored. So, this dissertation extends the investigation of Hoepner and Nilsson (2017) and focuses on the performance evaluation of US SRI corporate bond portfolios, between 2003 and 2018. The final dataset is free of survivorship bias and includes 8670 bonds, issued by 851 firms. The main goal of this investigation is to assess whether firms' ESG ratings impact the financial performance of the bonds issued by them.

The performance of SRI bond portfolios is first assessed using an unconditional model. Although this model has acceptable explanatory power of portfolio returns, it assumes risk and returns as stationary over time. Thus, to allow for variations in the state of the economy, further analysis is conducted using a conditional four-factor model. Evidence supports the use of conditional models, as there is strong evidence of time-varying betas and joint significance of alphas and betas across all approaches, as well as some evidence of time-varying alphas. High-rated, low-rated, and difference portfolios are constructed based on individual and aggregate ESG ratings. The analysis is conducted for positive and best-in-class screening investment strategies. Moreover, a set of robustness tests are applied, namely considering alternative cut-offs to construct the corporate bond portfolios, equally-weighted portfolios, and the exclusion of the financial sector. Finally, SRI bond portfolio performance is assessed over time.

The preliminary analysis from the unconditional model shows significant abnormal returns only for the low-rated portfolio based on social criteria (at the 5% level). Moreover, there is no evidence of abnormal alphas for any of the difference portfolios, in line with Halbritter and Dorfleitner (2015). Following, the regression outputs for the conditional multi-factor model show evidence of abnormal returns mostly for low-rated portfolios. Although more low-rated portfolios yield abnormal returns, in comparison to high-rated portfolios, the alphas for long-short portfolios are statistically equal to zero. Furthermore, the robustness tests show some differences in terms of alphas for high- and low-rated portfolios, although the overall conclusions still support that there are no differences between the performance of both. As cut-offs get more extreme, evidence of abnormal returns for high-rated portfolios appears, whereas evidence of abnormal alphas for low-rated portfolios disappear. Still, the difference remains insignificant. In terms of equally-weighted portfolios, there is an increase in alphas for both high- and low-rated portfolios, suggesting that smaller bonds tend to yield greater returns for investors. Finally, results for the exclusion of the financial sector show evidence of abnormal returns for alphas of both high- and low-rated portfolios, although the differences are, once again, not significant. These findings suggest that the inclusion of banks and financial institutions could be penalizing portfolio returns, especially portfolios composed of bonds issued by top-performing companies, as alphas of these portfolios gained statistical significance with this robustness test. In terms of differences between the performance of high- and low-rated portfolios, findings are in line with Pereira et al. (2019), and Halbritter and Dorfleitner (2015). Furthermore, Hoepner and Nilsson (2017) also report that difference portfolios yield alphas with no statistical significance, for aggregate ratings. However, results contrast with Polbennikov et al. (2017), who find evidence of incremental returns for high-rated portfolios, especially for portfolios based on governance ratings, and with Kempf and Osthoff (2007), who find long-short portfolios yield abnormal returns.

Regarding the best-in-class screening investment strategy, results are similar to the positive approach concerning high- and low-rated portfolios. However, there are some changes in terms of difference portfolios. Evidence suggests that portfolios composed of bonds issued by companies with poor CSR practices outperform portfolios of bonds of socially responsible firms for portfolios based on ESG combined screens. Nonetheless, this outperformance fades for the 10% and 5% cut-off rates, although there is still evidence of abnormal returns for high- and low-rated portfolios for both cut-off rates. Results contrast with Kempf and Osthoff (2007), who find that the best-in-class strategy yields the highest alphas for investors.

Finally, performance is evaluated over time. To do so, the sample period, which spans between 2003 and 2018, is divided into three subperiods (2003-2007, 2008-2012, and 2013-2018). Whereas no portfolios yield abnormal returns for the first subperiod, most high- and low-rated portfolios present positive and statistically significant returns, between 2008 and 2012. Again, there is no evidence supporting differences between the performance of those portfolios, except for portfolios based on the environmental and ESG combined ratings. For these portfolios, evidence suggests that, in the 2008-2012 period, low-rated portfolios outperform high-rated portfolios. However, this outperformance fades in the third subperiod. Alphas reached a peak in the second subperiod and decreased in the third subperiod, losing its statistical significance in most cases, as well. Furthermore, in the final subperiod, most evidence of abnormal returns disappeared, except for the high-rated portfolio based on environmental ratings and the low-rated portfolios based on social and ESG combined scores. These findings contrast with previous literature providing empirical evidence of the shunned-stock and errors-in-expectations hypotheses such as Derwall et al. (2011), Bebchuck et al. (2013), and Pereira et al. (2019).

In general, findings provide some evidence of positive and statistically significant abnormal returns for portfolios composed of bonds issued by high- and low-rated firms. Although the alphas for bond portfolios of bond-issuing companies with poor socially responsible behavior seem to be greater than those of more socially responsible corporations, for most approaches, the difference portfolios are statistically insignificant. Thus, although there is no evidence that investors can consistently achieve abnormal returns through SRI trading strategies, there is also no evidence that investors sacrifice financial performance by engaging in such strategies.

As expected, this research has some limitations. Some of these drawbacks are recurrent and, therefore, have been mentioned in previous literature. First, the investigation focuses only on the US market, and only bonds denominated in US$ and issued in the United States or international markets were considered. Although the US market is frequently used as a reference, conclusions from this study must be carefully drawn before extending them to different markets. Furthermore, the data provider may be considered a limitation as different data providers consider distinct criteria to construct ESG ratings. It would be interesting for upcoming literature to asses to which extent the inclusion of the financial sector impacts results, especially when extreme cut-offs or alternative approaches are applied. Finally, further literature could expand research to different geographic regions, and study the effects of considering different bond maturities and credit ratings.

References

Adcock, C. J., Cortez, M. C., Armada, M. R., & Silva, F. (2012). Time varying betas and the unconditional distribution of asset returns. *Quantitative Finance*, 12(6), 951–967.

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

Ayadi, M. A., & Kryzanowski, L. (2011). Fixed-income fund performance: Role of luck and ability in tail membership. *Journal of Empirical Finance*, 18(3), 379–392.

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.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1), 3-18.

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

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

Blake, C. R., Elton, E. J., & Gruber, M. J. (1993). The performance of bond mutual funds. *Journal of business*, 371-403.

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

Brzeszczyński, J., & McIntosh, G. (2013). Performance of Portfolios Composed of British SRI Stocks. *Journal of Business Ethics*, 120(3), 335–362.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), 57–82.

Christopherson, J. A., Ferson, W. E., & Glassman, D. A. (1998). Conditioning manager alphas on economic information: Another look at the persistence of performance. *Review of Financial Studies*, 11(1), 111–142.

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

Derwall, J., & Koedijk, K. (2009). Socially responsible fixed-income funds. *Journal of Business Finance and Accounting*, 36(1–2), 210–229. 58

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

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. *The Journal of Finance*, 50(4)(6), 1229–1256.

EUROSIF (2016). European SRI Study 2016. Retrieved June 26th 2019, from http://www.eurosif.org

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 Stock Returns. *The Journal of Finance*, 47(2), 427–465.

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. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1-22.

Ferson, W. E., & Schadt, R. W. (1996). Measuring Fund Strategy and Performance in Changing Economic Conditions. *The Journal of Finance*, 51(2), 425.

Filbeck, G., Gorman, R., & Zhao, X. (2009). The “best corporate citizens”: Are they good for their shareholders? *Financial Review*, 44(2), 239–262.

Freeman, R. (1984). E.(1984), Strategic management: A stakeholder approach. *Boston: Pitman*, *46*.

Gonenc, H., & Scholtens, B. (2017). Environmental and Financial Performance of Fossil Fuel Firms: A Closer Inspection of their Interaction. *Ecological Economics*, 132, 307–328.

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.

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

Hirshleifer, D. (2008). Psychological bias as a driver of financial regulation. *European Financial Management*, *14*(5), 856–874.

Hoepner, A. G., & Nilsson, M. (2017). No news is good news: Corporate social responsibility ratings and fixed income portfolios. *Available at SSRN 2943583*.

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 and 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(1), 15–36.

Ilmanen, A. (1995). Time-Varying Expected Returns in International Bond Markets. *The Journal of Finance*, 50(2), 481–506.

Jensen, M. (2001). Value maximisation, stakeholder theory, and the corporate objective function. *European financial management*, *7*(3), 297-317.

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

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. *Journal of Finance*, *72*(4), 1785–1824.

Markowitz, Harry, 1952, Portfolio selection, *Journal of Finance*, 7, 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.

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.

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

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 and Finance*, *48*, 180–193.

Oikonomou, I., Brooks, C., & Pavelin, S. (2014). The effects of corporate social performance on the cost of corporate debt and credit ratings. *Financial Review*, 49(1), 49–75.

Ortas, E., Moneva, J. M., & Salvador, M. (2014). Do social and environmental screens influence ethical portfolio performance? Evidence from Europe. *BRQ Business Research Quarterly*, 17(1), 11–21.

Pereira, P., Cortez, M. C., & Silva, F. (2019). Socially responsible investing and the performance of Eurozone corporate bond portfolios. *Corporate Social Responsibility and Environmental Management*, 26(6), 1407-1422.

Polbennikov, S., Desclée, A., Dynkin, L., & Maitra, A. (2016). ESG ratings and performance of corporate bonds. *Journal of Fixed Income*, 26(1), 21–41.

Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard business review*, 84(12), 78-92.

Refinitiv (2019). Thomson Reuters ESG Scores from Refinitiv. Retrieved April 22nd from https://refinitiv.com

Refinitiv (2020). Environmental, Social and Governance (ESG) Scores from Refinitiv. Retrieved on April 20th from https://refinitiv.com

Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking and Finance*, 32(9), 1723–1742.

Rudd, A. (1981). Social Responsibility and Portfolio Performance. *California Management Review*, 23(4), 55–61.

Scholtens, B. (2005). Style and Performance of Dutch Socially Responsible Investment Funds. *The Journal of Investing*, 14(1), 63–72.

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

Schwartz, M. S. (2003). The "Ethics" of Ethical Investing. *Journal of Business Ethics*, 43(3), 195–213.

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

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

Statman, M., & Shefrin, H. (1993). Ethiics , Fairess Efficiency and Markets in inancial and Meir Shefrin. *Financial Analysts Journal*, *49*(6), 21–29.

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

US SIF Foundation (2018) Report on US Sustainable, Responsible and Impact Investing Trends 2018. Retrieved may 24th 2020 from http://ussif.org/

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: journal of the Econometric Society*, 817-838.

Appendices

Appendix A. Number of companies and bonds by ICB Industry Groups

Industry Group	Companies	Bonds
Basic Materials	44	210
Consumer discretionary	133	765
Consumer Staples	46	472
Energy	72	465
Financials	161	3783
Health Care	83	475
Industrials	131	1192
Real Estate	45	252
Technology	72	433
Telecommunications	25	293
Utilities	39	330
	851	8670

Appendix B. Number of bonds by bond type

Bond Type	Observations
Straight	6601
Convertible	338
Floating	1483
Index-linked	96
Zero-coupon bond	152
	8670

Appendix C. Table of correlations between ESG ratings

	Environmental Score	Social Score	Governance Score	ESG combined Score
Environmental Score	1.000			
Social Score	0.702	1.000		
Governance Score	0.417	0.404	1.000	
ESG combined Score	0.878	0.863	0.716	1.000

Appendix D. Descriptive statistics on ESG scores for positive screening strategy portfolios – 10% cut-off rate

	Environmental Score		Social Score		Governance Score		ESG Score	
	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated	High-rated	Low-rated
Mean	95.33	27.27	91.33	30.77	90.36	34.08	88.31	35.49
Standard Deviation	3.49	6.66	3.53	6.47	3.38	10.14	2.63	6.69
Minimum	69.00	5.71	78.03	5.88	81.90	4.49	70.86	11.34
Maximum	99.09	42.07	99.04	45.48	98.79	48.65	97.66	46.83
Median	96.99	27.56	91.67	30.95	90.07	33.80	88.45	36.08
Skewness	-2.76	-0.51	-0.50	0.01	-0.36	-0.34	-1.01	-0.38
Kurtosis	18.42	2.82	2.97	2.74	2.38	2.46	7.59	2.54
Observations	56923	49705	58237	47133	54051	68665	50698	49681

Appendix E. Descriptive statistics on ESG scores for positive screening strategy portfolios – 5% cut-off rate

	Environmental Score		Social Score		Governance Score		ESG Score	
	High-rated	**Low-rated**	**High-rated**	**Low-rated**	**High-rated**	**Low-rated**	**High-rated**	**Low-rated**
Mean	96.27	22.26	92.20	26.63	91.94	28.50	88.92	30.21
Standard Deviation	2.41	4.91	3.33	4.67	2.58	9.58	2.38	4.67
Minimum	80.21	5.71	84.90	5.88	85.78	4.49	75.74	11.34
Maximum	99.09	32.01	99.04	37.28	98.79	40.60	97.66	38.12
Median	96.99	23.74	92.72	26.23	93.14	27.55	89.16	30.25
Skewness	-1.44	-0.83	-0.67	-0.32	-0.37	-0.21	-0.56	-0.65
Kurtosis	5.99	3.20	3.06	3.41	2.27	1.76	4.87	3.28
Observations	46538	25751	43078	25187	33469	32707	37826	24564

Appendix F. Table of correlations between risk factors

	Bond Market Factor	Default Spread Factor	Option Factor	Stock Market Factor
Bond Market Factor	1.000			
Default Spread Factor	0,293	1.000		
Option Factor	-0,308	0,626	1.000	
Stock Market Factor	0,303	0,704	0,328	1.000

Appendix G. Table of correlations between information variables in their detrended forms

	Term Spread (detrended)	Short-term Rate (detrended)	IRW (detrended)
Term Spread (detrended)	1.000		
Short-term Rate (detrended)	-0.733	1.000	
IRW (detrended)	-0.028	-0.268	1.000

Appendix H. Conditional performance evaluation model regression outputs for individual and ESG combined scores – performance over time

Panel A	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	**W1**	**W2**	**W3**
Env.Score									
High-rated	0.1164%	1.3609***	0.0130	0.0996	-0.0201	0.9437	0.0927	0.0602	0.0271
Low-rated	0.0208%	1.3870***	0.0028	0.1986	-0.0007	0.9179	0.1497	0.2219	0.0572
Difference	0.0956%	-0.0262	0.0103	-0.0990	-0.0194	0.1601	0.7301	0.3847	0.3406
Soc. Score									
High-rated	0.0555%	1.3272***	-0.0206	0.0858	-0.0007	0.9498	0.1263	0.1670	0.0573
Low-rated	0.0586%	1.4273***	0.0941	-0.0473	-0.0564	0.9243	0.4302	0.6281	0.5059
Difference	0.0031%	-0.1002	-0.1147*	0.1331	0.0556	0.0677	0.8281	0.4338	0.5231
Gov. Score									
High-rated	0,0950%	1.3714***	-0.0060	-0.0263	-0.0116	0.9389	0.3171	0.4109	0.3302
Low-rated	0,0537%	1.4353***	-0.0977	0.5297*	0.0035	0.8962	0.3827	0.8756	0.8422
Difference	0,0413%	-0.0638	0.0918*	-0.5560**	-0.0151	0.5634	0.7293	0.3026	0.4019
ESG Score									
High-rated	0.0873%	1.3599***	-0.0435	0.0773	-0.0034	0.9538	0.0584	0.0574	0.0171
Low-rated	0.0171%	1.3703***	0.0512	0.0604	-0.0186	0.8891	0.4668	0.8406	0.7160
Difference	0.0702%	-0.0104	-0.0947	0.0170	0.0152	0.1873	0.9526	0.1219	0.2038

Appendix H. (Continued)

Panel B	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W1	W2	W3
Env.Score									
High-rated	0.3694%	1.1944***	-0.1940	1.1641	-0.0087	0.8238	0.9717	0.0304	0.0522
Low-rated	0.7296%***	0.7333***	-0.0776	0.0668	0.1048	0.9051	0.2565	0.0000	0.0000
Difference	-0.3603%*	0.4611*	-0.1164	1.0973	-0.1136	0.3565	0.6991	0.0016	0.0021
Soc. Score									
High-rated	0.6626%**	0.8719***	0.0505	0.2704	0.0434	0.8388	0.0954	0.0061	0.0035
Low-rated	0.6807%***	0.8397***	-0.1216	0.0567	0.1083*	0.9284	0.2763	0.0000	0.0000
Difference	-0.0182%	0.0323	0.1721	0.2137	-0.0649	0.5043	0.0766	0.0005	0.0000
Gov. Score									
High-rated	0.4283%**	1.0867***	-0.1802	0.2408	0.0511	0.8614	0.1984	0.0332	0.0209
Low-rated	0.4096%**	1.0293***	-0.0936	0.4287	0.0465	0.9211	0.6547	0.0027	0.0011
Difference	0.0186%	0.0574	-0.0866	-0.1879	0.0046	0.6376	0.6092	0.1720	0.1421
ESG Score									
High-rated	0.3104%*	1.1072***	-0.1126	0.0640	0.0822	0.8607	0.3377	0.1456	0.1377
Low-rated	0.6875%***	0.7465***	-0.0665	-0.0623	0.1102	0.9242	0.2890	0.0000	0.0000
Difference	-0.3771%***	0.3607*	-0.0461	0.1263	-0.0281	0.5233	0.8654	0.0010	0.0005

Appendix H. (Continued)

Panel C	$\propto$	β_{Bond}	$\beta_{Default}$	β_{Option}	β_{Equity}	$Adj.R^2$	W1	W2	W3
Env.Score									
High-rated	0.0947%**	1.2975***	-0.1976***	0.1354	0.0757***	0.9624	0.3659	0.0000	0.0000
Low-rated	0.1021%	0.9483***	0.3530***	-0.5014**	0.0008	0.9344	0.3680	0.0000	0.0000
Difference	-0.0074%	0.3492***	-0.5506***	0.6367**	0.0749*	0.7042	0.7054	0.0023	0.0058
Soc. Score									
High-rated	0.0732%	1.2452***	-0.2232***	-0.0561	0.0379	0.9596	0.6361	0.0000	0.0001
Low-rated	0.1547%**	1.0077***	0.3560***	-0.4549**	-0.0192	0.9439	0.1755	0.0000	0.0000
Difference	-0.0816%	0.2375**	-0.5792***	0.3988	0.0571	0.8117	0.5500	0.0007	0.0022
Gov. Score									
High-rated	0.0441%	1.2156***	-0.0693	-0.1836	0.0071	0.9621	0.7711	0.0002	0.0006
Low-rated	0.0901%*	1.1421***	0.0784	-0.2197	0.0454*	0.9605	0.3991	0.0000	0.0000
Difference	-0.0460%	0.0734	-0.1477***	0.0361	-0.0383	0.6704	0.8623	0.0008	0.0027
ESG Score									
High-rated	0.0604%	1.2604***	-0.2214***	0.0040	0.0315	0.9565	0.5247	0.0000	0.0001
Low-rated	0.1256%**	0.9780***	0.3244***	-0.4764**	-0.0018	0.9408	0.3589	0.0000	0.0000
Difference	-0.0652%	0.2824***	-0.5458***	0.4804**	0.0333	0.8213	0.5290	0.0006	0.0018