MASTER THESIS

When does it pay to be Good?

Screening criteria and intensity of socially responsible investment funds and their financial performance

Student name: A.J. Zwijnenberg
Student Number: S1694545

Faculty: Behavioural, Management and Social Sciences
Study: Master Business Administration
Institution: University of Twente (NL)
Track: Financial management
Version: First Version

Place and date: Enschede, 26-02-2022

Supervisors: Dr. X. Huang
J. Osterrieder

UNIVERSITY OF TWENTE.

Abstract

This research aims to investigate the effect of screening intensity and screening criteria on the financial performance of socially responsible investment mutual funds. Screening intensity is the absolute number of individual criteria used in selecting investments, and screening criteria are used in the decision process for selecting investments based on specific ethical standards. The screening criteria are divided into six categories: Environment, Social, Governance, Product, and Shareholder Engagement. An Ordinary Least Square (OLS) regression is performed on a sample of 41 United States SRI mutual funds from the 1st of January 2015 till the 31st of December 2020. The main results calculate the risk-adjusted performance using the Fama-French five-factor model (2015). The results suggest that there are neither costs nor rewards to be gained when considering the number of screens used. The regressions on screening criteria show that the Environmental screen positively impacts the risk-adjusted return and that Shareholder engagement negatively affects the return. The two screening types are only significant when they are applied together. The Environmental screen's positive impact is more effective on the financial performance than the negative impact of being an involved shareholder. Therefore, screening for both criteria positively affects the performance.

Moreover, to provide rebuts results, this study provides additional results using the Capital Asset Pricing Model (CAPM), Fama-French three-factor model (1993), and Carhart four-factor model (1997) to calculate the risk-adjusted performance of SRI mutual funds. Both CAPM and Fama-French three-factor models (1993) show negative adjusted R squares, implying a very low or negligible explanation. If the Carhart four-factor model (1997) is used to calculate the risk-adjusted performance, the screening categories do not significantly impact the return of SRI funds. However, SRI funds' age and expense ratio significantly affect the performance when using the Carhart four-factor model (1997). To conclude, retail investors who want to invest in SRI mutual funds in the United States do not have to be concerned about sacrificing their financial gain to do good. They can even gain additional financial returns controlling for the five factors of Fama-French (2015) when investing in funds that screen for the environment and are engaged shareholders.

Keywords: Socially Responsible Investing (SRI), Investment Screens, Screening, Screening Intensity, Mutual funds, Sustainability, Environmental, Social, Governance, Products, Shareholder engagement, Fama-French, CAPM, Carhart four-factor model, risk-adjusted performance.

Table of content

1. Introduction

“Investors need to channel massive new volumes of capital into emerging markets over the coming decades in order to mitigate the droughts, famines and mass migrations that could be triggered by unchecked global warming” - Philipp Hildebrand, the vice chairman of BlackRock Inc (Tirone, 2021).[1]

The above citation from Hildebrand portrays the growing environmental concerns that lead to the inclusion of ethical behaviour into all aspects of life, including financial investing (Camilleri, 2020; Erragragui & Lagoarde-Segot, 2016). Globally, socially responsible investments (SRI) have experienced a substantial growth of 34% increase in two years, starting with $30.7 trillion at the beginning of 2018 (Global Sustainable Investment Alliance, 2018). SRI has grown from a niche investment strategy to a mainstream investment topic adopted by organisations and individuals (Erragragui & Lagoarde-Segot, 2016). An SRI investor aims to promote socially and environmentally sound corporate behaviour (Renneboog, Ter Horst, Zhang, 2008a). They avoid organisations that produce products that may cause health hazards or exploit people (Renneboog et al., 2008a). SRI investors expect organisations to focus on social welfare and value maximisation (Revelli & Viviani, 2015; Renneboog et al., 2008a). Additionally, SRI mutual funds screen their financial investments based on socially responsible criteria. Hence, their managers regularly screen their portfolios to evaluate their environmental, social, and governance qualifications (Camilleri, 2020).

Existing literature shows mixed results on the financial performance of SRI funds. For example, Renneboog, Ter Horst and Zhang (2008b) concluded that, on average, SRI funds in many European, North-American, and Asia-Pacific countries strongly underperform their domestic benchmark portfolios. Moreover, El Ghoul and Karoui (2017) found that funds that invest in more socially responsible firms exhibit weaker performance. The arguments against SRI mutual funds are that some appealing investment opportunities are excluded from the investment pool because they do not meet the ethical criteria (Barnett & Salomon, 2006; Renneboog et al., 2008b).

On the contrary, Cortez, Silva, and Areal (2009) concluded that European SRI funds give, in general, a performance that is comparable to that of conventional or SRI benchmarks. Additionally, they seem slightly higher when evaluating SRI funds’ performance assessments against socially responsible marks. The arguments supporting SRI funds are that social and environmental performance signals high managerial quality, translating into favourable financial functioning (Camilleri, 2020). Additionally, the stock markets might misprice corporate social responsible information in the short term, such that SRI funds may outperform conventional funds in the long run (Edmands, 2011; Revelli & Viviani,2015). Research reveals that the type of screening and the screening intensity matter for the financial return of a fund (Barnett & Salomon, 2006; Capelle-Blancard & Monjon, 2017; Renneboog et al., 2008b).

The literature refers to the screening intensity as the number of individual criteria used in selecting investments (Barnett & Salomon, 2006; Capelle-Blancard & Monjon, 2017; Lee et al., 2010; Renneboog et al., 2008b). Barnett and Salomon (2006) were the first to investigate the relationship between screening intensity and financial performance. Their study found a u-curve relationship between the number of criteria used and the financial performance of socially responsible investment funds. This relationship means that low levels of screening intensity provide funds with more opportunity for diversification of their portfolios, and hence have better financial performance. By contrast, funds with high screening intensity levels are better able to filter underperforming stocks from their portfolio. Barnett and Salomon (2006) claim that funds in the middle of screening intensity cannot properly diversify and do not achieve the rewards of screening intensely and subsequently

[1] https://www.bloomberg.com/news/articles/2021-10-07/blackrock-s-hildebrand-says-emerging-markets-need-climate-boost

underperform. Lee et al. (2010) found that the number of screens negatively impacts performance and lower systematic risk. However, Humphrey and Lee (2012) concluded that there is weak evidence between screening intensity and risk-adjusted performance in Australia.

A portion of the existing literature investigated if the type of screening applied by funds affects the performance. The screening used by SRI funds is a combination of negative and positive screening (De Colle & York, 2009; Renneboog, Ter Horst, Zhang, 2011). Negative screening excludes stocks that associate with specific sectors or products. Positive screening includes companies that meet ethical standards (De Colle & York, 2009). Barnett and Salomon (2006) found that screening based on community relations has a relatively more robust financial performance; however, screening based on equal employment opportunity and environmental performance may decrease the financial results. The researchers argued that the difference in the type of screening could be because investors rarely attend to information that extends beyond a 5-year horizon, and therefore, some of the financial benefits of specific social screens may not be visible until further in the future.

The overall literature is about the performance differences between SRI funds and conventional funds. Barnett and Salomon (2006) paved the way for researching the impact of the SRI screening process and its intensity on financial performance. However, much work is needed, especially now that SRI funds have become more mainstream (Erragragui & Lagoarde-Segot, 2016). The impact of different types of screening is essential for institutional and retail investors because having a better understanding of how SRI strategies may affect portfolio performance helps streamline the assets allocations and contest the preconception that taking into account social concerns is costly. Therefore, this research assesses the financial performance *within* the SRI mutual fund market. Specifically, the main subject is whether the financial performance of SRI funds is related to the characteristics of the screening criteria and the screening intensity.

This paper’s central question: *How do the screening intensity and criteria affect the financial performance of SRI funds?*

1.1 Scientific and Social Contribution

This paper will add to the existing literature on SRI mutual funds. First and foremost, this paper will use the relatively new Fama French five-factor asset pricing model (2015) to study the risk-adjusted return of the SRI funds. The five-factor model captures the size, value, profitability, and investment patterns in average stock returns and performs better than the Fama-French three-factor model (1993) (Blitz and Fabozzi 2017; Fama & French, 2015). The study adds value to the existing literature by using the Fama-French five-factor model (2015) to calculate the risk-adjusted return of SRI funds. To my knowledge, this study will be the first to use the five-factor model for calculating the return of SRI funds and use these outcomes to determine the effect of screening intensity and criteria on the performance of SRI funds.

Rather than only using one measure of performance, this study will also provide results using the Capital Asset Pricing Model (CAPM), Fama-French three-factor model (1993), and Carhart four-factor model (1997). Using multiple calculations for the performance provides insight for other researchers on the various types of risk-adjusted returns and the impact on the research findings. This insight enhances the discussion that the level of performance depends on the methodological choices made by researchers (Revelli & Viviani, 2015). Furthermore, this paper will advance the SRI debate by reviewing the significance of the SRI funds' screening intensity and individual criteria.

The growth in environmental awareness causes higher demand for more knowledge on the profitability of SRI funds. The social contribution of this paper is to supply more information on the rising demand for data on SRI investing. This growing demand is evident from a statement made by Vicotria Barron, head of sustainable investment for the BT Pension Scheme, in a Bloomberg news article:

"What's going to be a challenge is data. For us, it's really difficult as an international investor who essentially owns the whole market to get information across all of your asset classes." (Martin & Ritchie, 2021).[2]

Moreover, Philipp Hildebrand, the vice chairman of BlackRock Inc, stated, *"It is, in fact, only the beginning of one of the greatest structural changes the world has ever undergone. It means that investments that looked safe in the past could be existentially risky in the future."(*Martin & Ritchie, 2021*)*[3] The results in this paper will add knowledge to investors who want to invest in socially responsible mutual funds without decreasing their financial returns.

1.2 Thesis outline

This section portrays the outline of the thesis. The study continues with chapter two, the literature review that provides a theoretical background on mutual funds and SRI mutual funds. Chapter two will also present an overview of the arguments of the opponents and proponents of SRI fund performance with empirical evidence. The third chapter describes the difference in screening categories and concludes with various hypotheses. The following chapter describes the research methodology, and the measurements of the variables used, presenting empirical evidence of existing research methods to investigate the impact of screening criteria and intensity on the financial performance of SRI funds. Chapter five presents the data collection criteria used to gather the sample and the databases used to collect the data. The results and the robustness check are presented in chapter six. Afterwhich the conclusion of this study is stated. The final chapter presents the limitations and recommendations for future research.

[2] https://www.bloomberg.com/news/articles/2021-10-07/can-your-pension-resist-climate-change-u-k-funds-tally-risks

[3] https://www.bloomberg.com/news/articles/2021-10-07/can-your-pension-resist-climate-change-u-k-funds-tally-risks

2. Literature review

The literature review starts with general information on mutual funds and their characteristics, followed by a section that provides more detail on the background on SRI mutual funds. The last two parts present an overview of the arguments from proponents and critics of SRI mutual fund performance with empirical evidence.

2.1 Mutual Funds in General

Before specifying SRI mutual funds, this part discusses the general characteristics of mutual funds. A broad definition of an investment company is any arrangement in which several individuals invest in a company that invests in securities (Baumol, Goldfeld, Gordon & Koehn, 1990). Hence, a mutual fund is a financial vehicle consisting of a pool of money collected from individuals to invest in securities like stocks, bonds, and other assets (Baumol et al., 1990; Khorana, Servaes & Tufano, 2004; Pontiff, 1997). An open-end fund’s shares are directly purchased and sold from the fund at net asset value (NAV), ensuring that the price volatility and NAV volatility are equal (Pontiff, 1997). The shares of a closed-end fund are traded on stock exchanges; the market prices compute the value of the fund’s portfolio. The share price of a closed-end fund can differ from the fund’s NAV (Pontiff, 1997). Consequently, closed-end fund shares are sold at a discount to the NAV of their portfolio securities (Johnsen, 2003).

Mutual funds can offer different types of shares, also known as classes. Each class of shares from the same fund invests in the same investment portfolio; however, each class may have different fees and expenses[4]. Owning another type of the same fund will result in different investment returns. There are three main classes of shares, also known as A-shares, B-shares, and C-shares. The multiple share classes in a fund represent the fund manager’s several units to suit specific buyers.

There are a couple of ways mutual fund shareholders can realise a return on their investment. Investors can receive dividends earned on the investments (Baumol et al., 1990). Managers assign the fund’s assets according to a broad spectrum of allocation strategies and attempt to produce financial gains or income for the fund’s investors (Buldyrev, Flori & Pammolli, 2021). If the stakes in the fund’s portfolio increase in value, the NAV of a share will also increase, and an investor can realise a return by redeeming shares (Baumol et al., 1990; Johnsen, 2003; Pontiff, 1997).

2.1.1 Mutual funds characteristics

Fund characteristics can play an essential role in financial performance and strategy. For instance, the mutual fund industry is more prominent in countries with stronger laws, rules and regulations, specifically where investors’ rights are better protected (Khorana et al., 2004). Moreover, the industry is also more prominent in countries with a wealthier and more educated population (Khorana et al., 2004).

Funds characterise the industry of different sizes, whose portfolios may present either very concentrated investment positions or well-diversified portfolios (Kacperczyk, Sialm & Zheng, 2004). Moreover, funds can invest in organisations based on market capitalisation, where large-cap stocks have a market cap of $10 billion or more, and small-cap stocks generally have a market cap of $300 million to $2 billion. Kacperczyk et al. (2004) found that mutual funds that are more concentrated tend to follow distinct investment styles and that managers of more concentrated funds overweigh growth and small stocks. In contrast, managers of more diversified funds hold portfolios that closely resemble the total market portfolio. The researchers concluded that concentrated portfolios perform better than funds with a more diversified portfolio. Managers appear to remain focused on their few best bets and seek to scale up their existing investments when the net Asset Under Management (AUM) grows (Pollet & Wilson, 2008).

[4] https://www.investor.gov/introduction-investing/general-resources/news-alerts/alerts-bulletins/investor-bulletins-61

Pollet and Wilson (2008) somewhat contradict the findings of Kacperczyk et al. (2004). The researchers documented a positive relationship between diversification and performance (controlling for fund size and fund family size), where the connection between diversification and performance is significant for small-cap funds, presumable because they are more constrained.

The contradiction between the two researchers can be because of the variable industry size. This variable represents the fund's competition-size in the mutual fund industry. Hence, it is the size of the competition between mutual funds. Pástor, Stambaugh, and Taylor (2015) found a negative relationship between industry size and fund performance; this relationship is more significant for firms with higher turnover and volatility and small-cap funds. This relationship implies that the bigger the size of the competition within the mutual fund industry, the lower the fund performance will be. The researchers argue that these results appear sensible because funds that trade aggressively or trade illiquid assets discover that their higher trading costs achieve smaller profits when competing in a more crowded industry.

Ferreira, Keswani, Miguel, and Ramos (2013) enriched the literature by concluding that the fund size is only negatively related to fund performance in the United States. For non-US funds, size is associated with better performance. The researchers argue that the diminishing returns in the U.S. mutual fund industry are related to liquidity constraints faced by funds that have to invest in small and domestic stocks. They further state that international funds' performance does not depreciate with scale. Because international funds are less affected by the lack of new investment opportunities as the fund grows, they are not restricted to investing only in their local market. The researchers also acknowledged that other fund characteristics are essential in explaining a fund's performance. They listed that funds managed by large fund families display superior performance and that scale is not necessarily bad for performance in these cases.

There are also additional factors that influence the risk-adjusted returns of mutual funds. Carhart (1997) found that expense ratios, portfolio turnover, load fees, and transaction costs are significant and negatively related to a fund's performance. Just like any organisation, a mutual fund generates costs during its operations. Mutual funds can charge various fees to their investors to pay these costs. The fees charged by mutual funds are generally divided into two categories[5]. The first category is Shareholder fees: these costs occur with particular investor transactions and expenses linked to the fund's operating expenses. Several funds take in the costs associated with investors transactions by imposing fees directly on the investors. The other category is the annual fund operating expenses; funds pay these costs out of the fund's assets. The various fees based on the fund assets are, generally, a percentage of the fund's net assets (Baumol et al., 1990). Appendix I displays various fees listed on the official website of the U.S. government about investing.

Fees may differ between instruments in different countries. For example, Renneboog et al. (2008b) found that fees vary between countries and that the total fees are at their lowest in The Netherlands and Belgium and the highest in Malaysia. Addionally, Bauer et al. (2006) concluded that Australian domestic ethical funds have higher fees than conventional funds; however, they found the opposite to be true for international ethical funds.

Investors in mutual funds can select funds with all, some, or none of these fees. Hence, researchers are investigating if funds can justify these fees on a cost-benefit basis. Renneboog et al. (2008b) found that fund management fee significantly reduce the risk-adjusted returns of both SRI and conventional funds. Additionally, Dellva & Olson (1998) uncovered that 12b-1 fees, deferred sales charges, and back-end fees increase expense, and only a limited number of funds with these fees earn a risk-adjusted return that can justify these fees. Nevertheless, the absence of these fees does not mean superior performance since most funds without any fees also make a negative risk-adjusted return (Dellva & Olsen, 1998).

[5] https://www.investor.gov/introduction-investing/investing-basics/glossary/mutual-fund-fees-and-expenses

2.3 Background on SRIs

Socially responsible investments (SRI) is an investment process that incorporates social, environmental, and ethical factors into the investment decision-making process (Renneboog et al. 2008a). Haigh and Hazelton (2004) state a straightforward definition for the procedure of SRI:

"Socially responsible investment (SRI) is a term used to refer to the practice of directing investment funds in ways that combine investors' financial objectives with their commitment to social concerns such as social justice, economic development, peace or a healthy environment." (Haigh & Hazelton, 2004, p. 59)

SRI funds aim to provide investors with higher returns by investing in corporations that excel in corporate social responsibility. The screening criteria funds use to compose their portfolio vary widely between funds (Capelle-Blancard & Monjon, 2012). They range from simple screens, like excluding specific industries, to screens related to governance issues and the well-being of employees (Renneboog et al., 2008a). The origin of SRI investing can be traced back to ancient times when religious principles were directing the investment decisions (Renneboog et al., 2008a). Modern SRI focuses on individual investors' varying ethical and social convictions (Renneboog et al. 2008a). The current type of ethical investing can be traced back to 1971 when investors opposed the Vietnam War (Renneboog et al., 2008a). The SRI industry has experienced high growth during the last decade and has become a fashionable investment method (Lagoarde-Segot, 2016).

The increase in SRI attracted the attention of various scholars, which provided valuable literature on SRI. The primary focus is on the performance of SRI funds, pursuing to understand whether this type of investment has abnormal returns. Several empirical studies have attempted to establish a link between the effect of introducing non-financial criteria in the investment process and the financial performance of SRI funds.

A meta-analysis done by Revelli and Viviani (2015) looked at studies that compared the performance across SRI and non-SRI groups. They included 85 empirical studies involving 190 experiments from 1972 to 2012. The researchers concluded that globally there is no actual cost or benefit to investing in SRI. However, the level of performance depends on the methodological choices made by researchers. Jones, van der Laan, Frost and Loftus (2007) tried to address the difference in methodology by investigating the returns performance of 89 ethical funds in Australia over 1986-to 2005. Many previous studies do not control for size effects with SRI funds; Jones et al. (2007) used a multifactor model containing size, book-to-market, and momentum factors. They found that SRI funds significantly underperform the market in Australia. However, they stated that the energy sector promptly pushed the Australian stock market performance in recent years. Many SRI funds avoid these sectors; hence, investment strategy and portfolio screening may be essential factors driving financial performance.

Renneboog et al. (2008b) concluded a difference between SRI funds' performance and location. They found that SRI funds in the United States, the United Kingdom, and other European and Asia-Pacific countries underperform their domestic benchmarks. However, some countries such as France, Japan, and Sweden were exceptions. The risk-adjusted return of SRI funds in these countries are not significantly different from the performance of conventional funds. Capelle-Blancard and Monjon (2012) emphasise the various screening practices used in countries. The researchers stated that negative screening is widespread in the United States, while continental Europe's best-in-class approach is the standard, these different screening types can have mixed performance results. Auer (2016) concluded that a simple negative screening method that excludes unrated stocks provides significantly higher performance than passive benchmark strategies. They also concluded that investors could choose SRI consistent with their values and beliefs without sacrificing performance. They can even come up with higher performance. However, these arguments only hold when investors apply a particular ethical screening strategy.

The screening intensity of SRI funds is also an essential topic for the financial return of the fund. There is evidence of a curvilinear relationship between screening intensity and financial return (Barnett & Salomon, 2006). This relationship implies that as the number of screens used by SRI funds increases, financial returns decline at first but then rebound as the number of screens reaches a maximum. Some studies found that higher screening intensity reduces the risk-adjusted return (Capelle-Blancard & Monjon, 2014; Renneboog et al., 2008b). However, these results only hold for sector-specific screening criteria (Capelle-Blancard & Monjon, 2014).

2.4 The underperformance of SRI funds

This section is divided into two parts; the first part discusses the theoretical foundation for the underperformance of SRI funds. The second part will provide empirical evidence for the criticism against SRI funds.

2.4.1 Theoretical foundation on the underperformance of SRI funds

Different theories support the argumentation of the underperformance of SRI funds. The first theory for the underperformance of SRI funds is the Modern Portfolio Theory (MPT) (Markowitz, 1952). MPT theorises that investors can construct a portfolio of multiple assets that will result in greater returns without a higher level of risk. By creating a diversified portfolio, investors can maximise their returns without unacceptable levels of risk. The theory distinguishes two types of risk: systematic and unsystematic risk. Systematic risk is inherent in the market's volatility, called market risk (Brealey et al., 2019). Unsystematic risk, also called specific risk, is associated with individual security volatility (Brealey et al., 2019). Investors may assemble their portfolios so that another security within the portfolio offsets the unsystematic risk. The offsetting of unsystematic risk is called diversification, and it can diversify the unsystematic risk away (Brealey et al., 2019). Efficient working capital markets will reward investors for bearing systematic risk. However, holding unsystematic risks is not rewarded by the market (Barnett & Salomon, 2006). When an investor carries unsystematic risk, it fails to reach the efficient frontier, wherein the risk/return trade-off is optimised (Barnett & Salomon, 2006; Brealey et al., 2019). The challengers of SRI funds argue that by applying screening criteria, investors limit the full diversification potential of their portfolios and carry excessive unsystematic risk, which diminishes the return (Brealey, Myers & Allen, 2019; Markowitz, 1952).

Another popular explanation for the underperformance of SRI stocks is the shunned-stock hypothesis (Derwall et al., 2011). A group of investors shuns certain "sinful" stocks (i.e. stocks of companies that profit from tobacco, alcohol, and gambling); when this group becomes large enough, the price of sinful stocks will deteriorate as prices fall and expected returns increase. Consequently, investors who include sinful stocks in their portfolios will gain abnormal returns (Derwall et al., 2011). Additionally, sin stocks conflict with societal norms; consequently, institutional investors who are vulnerable to public opinions, such as public pension funds, avoid these stocks (Hong & Kacperczyk, 2009). By not buying sin stocks, these norm-constrained investors cause the sinful stocks to be relatively cheaper, all else being equal, and higher expected returns.

An Additional theory that could explain the underperformance of SRI is the theory of SRI costs (Capelle-Blancard & Monjon, 2012). Opponents of SRI point out that any effort regarding the social responsibility of firms is costly and will result in above-average costs that subsequently will cause below-average financial performance (Renneboog et al., 2008a; Revelli & Viviani, 2015). Revelli and Viviani (2015) categorised SRI's expenses into two parts. The first category of additional costs for SRI funds is the expenses associated with determining which stock belongs to the SRI universe, the universe selection costs. This universe selection costs is because reporting on social and ethical aspects is not as organised and standard as economic factors, relying on accounting procedures. Therefore, determining which stock belongs to the SRI universe is more time-consuming. Secondly are the costs of active management. SRI portfolios are relatively small; therefore, the expense generated by active managers is proportionately higher than conventional funds.

2.4.2 Empirical results on the underperformance of SRI funds.

Lee et al. (2010) investigated 61 U.S. equity funds to explore if screening intensity affects the fund's systematic risk and return. They found that highly-screened funds suffer in terms of risk-adjusted performance. The performance of a fund decreases by 70 basis points per screen applied. Likewise, Renneboog et al. (2008b) investigated 440 live and dead equity SRI mutual funds and 16.036 live and dead equity conventional mutual funds from 17 countries. The researchers concluded that SRI funds have significantly negative alphas and are underperforming compared to their traditional counterparts. They concluded that investors pay the price for ethics due to their aversion to unethical behaviour.

Additionally, Derwall et al. (2011) found evidence for the shunned-stock hypothesis. The researchers concluded that the group of ethical investors is large enough to affect the supply and demand for securities. Stocks that ethical investors avoid by using screening criteria earn abnormal positive returns. Donath et al. (2018) focused on the United States market and utilised the Markowitz and Sharpe market models to determine the market value of SRI and non-SRI mutual funds. They revealed that conventional funds show better performance to yield higher returns. Lastly, Capelle-Blancard and Monjon (2014) explored if the financial performance of SRI mutual funds is related to the features of the screening process based on French SRI funds. They concluded that higher screening intensity reduces the risk-adjusted return. Though, the results hold only for sector-specific screening criteria.

2.5 Out-performance of SRI funds

This section discusses the theoretical foundation for the out-performance of SRI funds and provides empirical evidence for the performance of SRI funds.

2.5.1 Theoretical foundation on the out-performance of SRI funds

The supporters of SRI funds commonly use the stakeholder theory to support their arguments for the performance of SRI funds. The stakeholder theory describes businesses as relationships among groups with a stake in an organisation's activities (Jones, Harrison & Felps, 2018). Therefore, firms need to build strong positive relations with their stakeholders to perform better (Barnett & Salomon, 2006). Consequently, SRI proponents argue that because stakeholder relationships matter to financial performance, social responsibility is not purely a cost but a strategic investment (Jones et al., 2018). A close relationship between an organisation and stakeholders can lead to higher levels of joint value creation because of more efficient joint coordination, knowledge sharing advantages, higher-quality stakeholders, reduced transaction costs, and greater moral motivation. These benefits of a close relationship can exceed the costs of a strategy used to develop and maintain it (Jones et al., 2018). Additionally, organisations that successfully create this relationship may enjoy a sustainable competitive advantage since such benefits are rare and difficult to imitate (Jones et al., 2018).

Even if social screens may limit the investment pool of funds, advocates of SRI argue that managers might have better odds of avoiding stocks that would perform poorly in the future by being more critical of their investments (Barnett & Salomon, 2006). The logic behind this is that a better stock selection might offset the lack of diversification among stocks.

Derwall et al. (2011) offer a counterpart to the shunned-stock hypothesis. They argue that corporate social responsibility (CSR) information is value-relevant, and the financial markets do not understand it well enough to predict the value is the fundamental nature of the 'error-in-expectations hypothesis'. For this hypothesis to hold, the stock price should not reflect the value-relevant information related to CSR practices. Derwall et al. (2011) rationalise that the market fails to value CSR practices properly. Since CSR is a multidimensional and partially subjective concept, investors lack the tools to sufficiently measure CSR practices and their effect on the organisation's value.

2.5.2 Empirical results on the out-performance of SRI funds

Cortez et al. (2008) investigated the performance of SRI mutual funds from seven European countries that invested globally and (or) in the European market. They found that European SRI funds generally produce neutral performance compared to conventional and socially responsible benchmarks. The performance is slightly higher when funds are evaluated with socially responsible indices. They found that investors who wish to hold European funds can add social screens to their investment decisions without compromising financial performance.

Moreover, Kacperczyk et al. (2004) investigated the relationship between the industry concentration and the performance of actively managed United States mutual funds from 1984 to 1999. The researchers found that, on average, more concentrated portfolios perform better than more diversified funds after controlling for risk and style differences. Managers of more concentrated funds overweigh growth and small stocks, whereas managers of more diversified funds hold portfolios that closely resemble the total market portfolio. Their results indicate that SRI funds do not have to be diversified to do well and that the investment ability of managers is more important than being diversified. Complementary, a study done by Erragragui and Lagoarde-Segot (2016) found that the difference in return between SRI funds and conventional funds is insignificant. They argued that SRI funds have become mainstream and can diversify better.

Lastly, Blitz and Fabozzi (2017) explored the performance of sin stocks using global data until the end of 2016. The researchers concluded that sin stocks exhibit a significant positive CAPM alpha. However, this alpha disappears when controlling for the classic factors such as size, value, momentum and exposure to the two new Fama and French (2015) factors: profitability and investment. Therefore, the researchers argue that SRI investors can restore their portfolios’ expected returns by ensuring that their factor exposures do not deteriorate when excluding sinful stock.

3. Screening Criteria and Hypothesis

Different screening criteria provide investors with various mutual funds that best fit their values and beliefs. However, providing a clear-cut definition of screening criteria remains a difficult task. The forum for sustainable and responsible investment (USSIF) provides an overview of SRI fund managers' screens constructed by Bloomberg; this definition is primarily used in the United States[6]. The website offers six categories with subcriteria: environmental, Social, Corporate governance, Products, Other qualitative criteria, and Shareholder engagement [7].

The European equivalent, Eurosif, provides a different set of screening methods. Eurosif articulates seven general strategies rather than screening categories[8]. Appendix II provides a summary of the screening strategies used by Eurosif. These strategies evolved from a risk management focus to seeking opportunities to generate social and environmental impact alongside a financial return, with a financially sustainable long-term mindset (Eurosif, 2018).

The screening method used by USSIF is used in this research; an in-depth discussion on the screening method is presented later in this chapter. The difference between USSIF and Eurosif in the screening method signals that a universal definition may be difficult to formulate. Additionally, the markets are dynamic and innovative by nature. Therefore, it can be challenging to identify such a uniform set of screens as they continually change.

Regardless of the different definitions, researchers have established that SRI funds construct their portfolios with negative and positive screens (De Colle & York, 2009; Renneboog, Ter Horst, Zhang, 2011). Negative screening is the exclusion of stocks from a specific sector or product. For example, funds can exclude organisations that partake in the defence industry. (De Colle & York, 2009). Positive screening includes companies that meet superior standards on SRI issues (De Colle & York, 2009).

Funds often combine positive screens with a '*best in class'* approach (Renneboog et al., 2008a; Renneboog et al., 2008b). The best-in-class method is one of the Eurosif strategies; investors rank firms within each industry or market sector based on CSR criteria. Once all the firms have been categorised, only those that pass a minimum threshold are included in the portfolio (Renneboog et al., 2008a; Renneboog et al., 2008b). Kempf and Osthoff (2007) tested this trading strategy by buying stocks with high CSR ratings and selling stocks with low CSR ratings. They concluded that investors could earn remarkable high abnormal returns by following the simple long-short strategy. They stated that investors could achieve these returns by applying the positive screening approach or the best-in-class approach; however, investors cannot achieve these results using a negative screening approach.

Eurosif mentioned three other strategies for SRI investors (Eurosif, 2018). First, *impact investing* is investments made into organisations that have the intention to generate social and environmental impact. Eurosif stated in their 2018 brochure that impacts investing was the fasted growing strategy. These investments are often project-specific and expect a positive financial return. Secondly, *stewardship and engagement* are engagement activities and active ownership through voting shares and engagement with companies on ESG matters. This strategy is a long-term process seeking to influence behaviour or increase disclosure. The third strategy is *sustainability-themed*. This type of strategy is focused on investing in any effort related to sustainable development. Eurosif documented the growth of 146% between 2013-and 2015 for sustainability-themed investment. This strategy focuses on long-term investments associated with sustainable development, such as energy efficiency, carbon reduction, and water scarcity. The following section will discuss the different screening methods used by USSIF and articulate corresponding hypotheses.

[6] https://www.ussif.org/about
[7] https://www.ussif.org/policy
[8] http://www.eurosif.org/policy/

3.1 Screening intensity

The first hypothesis focuses on the screening intensity of a portfolio and the financial performance of SRI funds. Screening intensity refers to the number of screens that SRI mutual funds apply. This intensity is measured as the absolute number of screens used (Barnett & Salomon, 2006; Renneboog et al., 2008b; Lee et al., 2010). The greater an SRI fund's screening intensity, the smaller its pool of potential investments. On the contrary, smaller screening intensity implies a more extensive collection from which a fund manager might select investments and a closer resemblance to a broadly diversified portfolio. Barnett and Salomon (2006) combine the modern portfolio theory and stakeholder theory into a u-shape relationship between the screening intensity and financial performance. The logic behind their hypothesis is that the lack of diversification might offset higher screening intensity with a better stock selection. The funds in the middle of the u-shape are deemed as 'stuck in the middle by Barnett and Salomon (2006). These funds apply too many screens to diversify their unsystematic risk effectively, yet they are not using enough screens to eliminate underperforming firms from their portfolio. A study done by Erragragui and Lagoarde-Segot (2016) found that the difference in return between SRI funds and conventional funds is insignificant because they argued that the increased pressure on organisations to concern themselves with CSR issues had improved the investment pool for SRI funds. They imply that nowadays, SRI fund managers have more organisations in their investment pool to choose from; thereby, they are more capable of diversifying their portfolios.

For this paper, the hypothesis is formulated to see if the u-shape hypothesised by Barnett and Salomon (2006) is still relevant now that socially responsible behaviour has become more mainstream. Theorising that the 'stuck in the middle' funds can now diversify their portfolio better and, therefore, may not carry unsystematic risk. Accordingly, the first hypothesis:

Hypothesis 1: The relationship between SRI funds' intensive screening and financial performance is not u-shaped.

3.2 Environmental screens

The first type of USSIF screening category is that of environmental screens. At the start of 2020, investment managers consider environmental criteria across nearly $16 trillion in AUM[9]. USSIF categorise the environmental screen into three sets. The first screen is *Climate/Clean Tech*, which looks at the risk and opportunities associated with climate change, greenhouse gas emissions and sustainability of production. The second one is *Pollution/ Toxics*, which considers business toxicity and how organisations manage pollution and waste. This group includes recycling, waste management and water purification. The third category, *Environment/others*, focuses on environmental issues that the previous groups do not capture.

Nowadays, organisations with poor ecological performance risk consumer disfavour, protests by activist groups, negative media coverage, and general reputation loss (Ambec & Lanoie, 2008; Barnett & Salomon, 2006). Better environmental understanding can increase revenue because organisations have better access to specific markets; they differentiate products and sell pollution-control technology (Ambec & Lanoie, 2008). Additionally, solid environmental performance can lead to the opportunity to reduce risk management and relations with external stakeholders, cost of materials, energy and services, cost of capital, and lastly, cost of labour (Ambec & Lanoie, 2008).

However, Barnett and Salomon (2006) concluded that environmental screening standards negatively influence the financial benefits. The researchers argue that environmental screening benefits are long-term and therefore not considered by investors since they rarely attend to information beyond 5-years. Moreover, Auer (2014) found that environmental screens do not significantly add value. It might be that social investors are actively seeking these types of investment that the increase

[9] https://www.ussif.org/environmental

in demand is lowering the expected returns (Derwall et al., 2011). Therefore, I argue that the effects cancel each other out:

Hypothesis 2: Screening based on environmental criteria will not earn higher financial returns than those not.

3.3 Social screens

The main category of social screening contains four subdivisions. The first group is *Community Development*, which focuses on supporting low- and medium-income communities. For example, fair consumer lending and affordable housing. Secondly, *Diversity & Equal Employment Opportunites* considers the diversity and equal employment opportunity policies and practices relating to employees, company ownership or contractors. Following is the group of *Human Rights* that assesses the risk associated with human rights and companies' respect for human rights within their interval operations and the countries in which they operate. With particular emphasis on relations with indigenous people, supply-chain management and conflict zones. The fourth group is about *Labour Relations*; it considers how the organisations treat their employees. For example, it looks at health and safety, employment and retirement benefits, union relations and employee involvement. Lastly, *Conflict Risk* is the exclusion or partial exclusion of companies that conduct business in countries identified as repressive regimes or state sponsors of terrorism.

Employees are an essential factor in the success of an organisation (Edmans, 2011). Since organisations rely more on the employees' creativity and knowledge to create value, labour relationships have become increasingly important. For example, Edmans (2011) found that high levels of employee satisfaction generate exceptional long-horizon returns. The auteur further argues that SRI screens based on employee welfare may improve investment performance, in contrast to current views that any SRI screen lowers investors' returns. Additionally, organisations that take care of local communities will reap many returns, including better schools, fewer local restrictions, and a better infrastructure to support the firm (Waddock & Smith, 2000). Favourable local relations can also decrease the likelihood and intensity of protests. Moreover, this beneficial relationship can increase the probability of successful bargaining with local governments (Waddock & Smith, 2000).

However, Chapple and Humphrey (2013) could not find a difference in the performance of gender-diverse and all-male board portfolios. They found weak evidence of a negative correlation between having a diversified board and performance; however, diversity is positively related to performance in some industries. A study done by Shakil (2021) concluded that board gender diversity is crucial to reducing firms' financial risk. They looked at the board members in the oil and gas industry and concluded that, in general, the board with a mix of male and female members faces less risk than a male-dominated board.

Scholtens (2008) found a significant positive relationship between financial and social performance. However, they emphasise that the different themes such as community involvement, employee relations, diversity, environment, and product do not all have the same type of interaction with financial return and risk. Additionally, he concluded that specific themes could cause a decrease in results. A study done by Cheng, Loannou and Serafeim (2013) concluded that firms with better CSR performance are associated with superior stakeholder engagement, which significantly reduces the likelihood of opportunistic behaviour and introduces a more efficient form of contracting with critical constituents. They stated that exceptional stakeholder relationships reduce potential agency costs and enhance profit through higher-quality relationships with customers, business partners, and employees.

Therefore, arguing that the positive effects will outway the negative impact, the hypothesis will be:

Hypothesis 3: Screening based on social criteria will earn SRI funds a higher financial return than those not.

3.4 Governance screens

The governance category composes of board issues and executive pay. *Board issues* consider the directors' independence, diversity, compensation and responsiveness to shareholders. *Executive Pay* deals with whether pay policies are reasonable and aligned with shareholders' and stakeholders' long-term interests.

There is a general understanding that good governance is essential for protecting long-term shareholder value, decreasing the likelihood of committing fraud (Farbar, 2005). At the beginning of 2020, governance categories, including shareholder rights, executive pay, transparency, and anti-corruption, were applied to nearly $16 trillion in assets under management[10]. Incentives provided to top management is an essential fundamental attribute of the corporate governance system (Eccles, Ioannou, & Serafeim, 2014). Organisations that find ESG important are more likely to align the executive incentives to environmental, social, and external non-financial goals (Eccles et al., 2014). Moreover, organisations with more substantial shareholder rights have higher firm value, higher profits, higher sales growth, and lower capital expenditures (Gompers, Ishii, & Metric, 2003). Moreover, Auer (2014) stated that the most important was the governance screen to achieve higher performance out of the three screening types. Therefore, the following hypothesis is:

Hypothesis 4: Screening based on governance criteria will earn SRI funds a higher financial return than those not.

3.5 Products screen

The product category of USSIF is concerned with the exclusion of product types: *Alcohol, Animal cruelty, Defense/Weapons, Gambling, and Tobacco*. Funds can benefit from excluding particular industries from their portfolios. When companies are convicted of misdeed, the market may interpret that as an indicator that the firm will incur more costs from penalties, or from stakeholders less willing to cooperate, or will suffer lower revenues due to reduced future demand for the firm's product and services resulting from the company's damaged reputation (Margolis, Elfenbein, & Walsh, 2007). As mentioned above, the shunned-stock hypothesis argues that these products will generate abnormal returns since demand detergents, prices fall and expected returns increase (Derwall et al., 2011). However, Blitz and Fabozzi (2017) claimed that investors could counter this abnormal return by sin stock by exposing their portfolio's to the same factors to restore their expected return. Therefore, this study argues that fund managers can reveal their portfolio's to the same factors as sinful stocks to regain their expected return. Hypothesising that the effect will cancel each other out:

Hypothesis 5: Screening based on products criteria will not earn SRI funds a higher financial return than those that do not.

[10] https://www.ussif.org/governance

3.6 Other / Qualitative

Other/ Qualitative screens consider other environmental, social, governance, or product-specific criteria not mentioned by the different screens. This research adds other qualitative screens to the model. However, this study does not state a hypothesis based on this category since the incorporated details vary between the funds.

3.7 Shareholder Engagement

The final category is that of Shareholder Engagement which concentrates on the fund's effort in filing or co-filing shareholder resolution and (or) the fund's engagement in private dialogue on ESG issues with companies in the investment portfolio. Alda (2019) researched the influence of S.R. pension funds as institutional shareholders on the sustainable development of investee firms. The researcher found that pension funds' long-term nature provides stability and helps develop long-term ESG practices. This behaviour increases the transparency towards the shareholders and stakeholders, improving market efficiency. However, the ESG practices promoted by the pension funds do not affect short-term profit. Haigh and Hazelton (2004) argue that currently, SRI funds lack the power to create significant corporate changes and that shareholder advocacy has been largely unsuccessful. Contrary, Uysal, Yang and Taylor (2017) found that shareholder activism on social issues can yield significant outcomes if done correctly. The way activist advocates and frame an issue affects how firms will respond to the issue. Done correctly, and the firm will see the issue as an opportunity. Done wrongly, the firm will see the issue as a threat. Hypothesising that the effect will cancel each other out:

Hypothesis 6: Screening based on Shareholder Engagement will not earn SRI funds a higher financial return than those that do not.

4. Methodology

The focus of this chapter is on the methodology of this research. Scholars have used different techniques to investigate the financial performance of SRI funds. This study explores the effects of SRI funds' screening category and intensity on financial performance. This section will briefly explain the research methods used in previous literature, which establishes the method used in this study. Moreover, this section defines and describes all the variables used in this research. The final part of this chapter provides an overview of the statistical method used for analysing the data.

4.1 Research methods

As stated, this paper aims to investigate the effects of screening criteria and intensity on the financial performance of SRI funds. Several studies also have moved away from comparing the performance of SRI funds with those of conventional funds to investigate the performance within the SRI funds industry. The findings of previous papers tend to unite, though neither the data nor the econometric specifications are perfectly comparable. Barnett and Salomon (2006) and Lee et al. (2010) examine, respectively, 67 and 61 United States SRI funds over a longer time of 1972 till 2000 and 1989 to 2006. In contrast, Renneboog et al. (2008b) examine a large sample of 440 live and dead equity mutual funds worldwide from 1991 to 2003. Humphrey and Lee (2012) concentrated on Australian SRI funds from 1996 to 2008. The overall methodology is somewhat the same in all the studies.

The ordinary least squares (OLS) regression method is the standard regression analysis in this field of research (Capelle-Blankcard & Monjon, 2012; Barnett & Salomon, 2006; Lee et al., 2010; Renneboog et al., 2008b). OLS regression is a linear least-squares method; linear models predict values that fall in a straight line by having a constant unit change (the slope) of the dependent variable for a constant unit change of the independent variable (Hair, Black, Babin & Anderson, 2014). The principle of least squares is to minimise the sum of the squares of the differences between the observed dependent variable and those predicted by the linear function of the independent variables (Hair et al., 2014). To test all the hypotheses, the OLS regression, as most pronounced in the existing empirical studies, is used in this study.

There needs to be a significance level in regression analysis to determine if the results reject or support the hypothesis. The most widely used significance level is 0.05, although researchers use levels ranging from 0.01 to 0.10 (Hair et al., 2014). This study will also apply a significance level of 0.05. meaning that if the p-value is lower than 0.05, the coefficient is a meaningful addition to the model because changes in the predictor value will change the response variable.

4.2 Measurement of variables

There are three sets of variables for the empirical model of this research, namely, dependent, independent and control variables. The variables selected for this research align with previous literature to enhance comparability. Appendix IV provides an overview of all the variables used in this research.

4.2.1 Risk-adjusted performance

The dependent variable used in this research will be the risk-adjusted performance of a given SRI fund. The risk-adjusted performance (RAP), also called the risk-adjusted return, measures an investment's return after considering the degree of risk taken to achieve it. For this study, the RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end on a 12-month estimate, adjusted by specific risk factors.

The literature's primary model used for the risk-adjusted performance is CAPM (Barnett & Salomon, 2006; Capelle-Blankcard & Monjon, 2012). CAPM explains the stock return as a function of the market return. The Fama-French three-factor model (1993) is also frequently used to calculate risk-adjusted performance (Capelle-Blankcard & Monjon, 2012; Lee et al., 2010; Renneboog et al., 2008b). This model describes stock return through three factors: market risk, the outperformance of small-cap companies relative to large-cap, and the outperformance of high book-to-market value companies versus low book-to-market value organisations. Additionally, the Carhart four-factor (1997) model appears in the literature for calculating the RAP (Capelle-Blankcard & Monjon, 2012; Lee et al., 2010; Renneboog et al., 2008b). This four-factor model adds a factor to the Fama-French three-factor model (1993), namely, the momentum. Momentum is the average return on equal-weighted high return portfolios minus the average return equal-weighted on low return portfolios (Carhart, 1997). In other words, momentum is the speed or velocity of price changes in stock, security, or tradable instruments. Henceforward, the Fama-French Three-factor (1993) will be noted as FF3 and the Carhart four-factor model (1997) as CH4.

In 2015, Fama and French improved their model by adding more factors. The researchers concluded that the Fama-French Five-factor model (2015) (FF5) captures the size, value, profitability, and investment patterns in average stock return better than the FF3. Chiah, Chai, Zhong, and Li (2016) investigated if the FF5 outperformed the FF3 in Australia. The researchers found that the five-factor better explains the asset pricing variances. Furthermore, Erding (2017) compared the CAPM, FF3 and the FF5 for the Turkish stock market and concluded that the FF5 better explains the typical variation in stock return than the other models. For this research, the FF5 will be the primary source to calculate the RAP for the SRI funds:

$$r_{i,t} - r_{f,i} = \alpha_{5f,i} + \beta_{m,i}\left(r_t^m - r_t^f\right) + \beta_{smb,i} * r_t^{smb} + \beta_{hml,i} * r_t^{hml} + \beta_{rmw,i} * r_t^{rmw} + \beta_{cma,i} * r_t^{cma} + \varepsilon_{i,t} \quad (1)$$

$\alpha_{5f,i}$ alpha resembles the interception of the return beyond what is predicted by the corresponding model; these are monthly alpha's estimated over a 12-month period, also defined as the RAP. r_t^m symbolises the return on the market portfolio for month t, r_t^f is the risk-free rate of return on a one-month United States T-bill. $\beta_{m,i}$ captures the fixed beta of fund i. $\varepsilon_{i,t}$ is the random error. r_t^{smb} is the return of the small-minus-big (SMB) factor that is proxied by the return spread between a small-cap portfolio and a large gap portfolio. Equivalent, r_t^{hml} is the return of the high-minus-low (HML) factor, the gap between a high book-to-market ratio and a low book-to-market ratio portfolio. RMW stands for the robust minus weak where r_t^{rmw} is the average return of robust operating profitability portfolios minus the average return on the weak operating profitability portfolios. r_t^{cma} is the average return of conservative investment portfolios minus the average return on aggressive investment portfolios. The $r_{i,t}$ denotes the return of fund i in months t, and it will be computed with calculation 2. The return will be based on monthly data because monthly data is readily available from the database since the fund's interception. In this paper, the monthly return will be calculated as:

$$r_{i,t} = \frac{NAV_{f,t} + D_t}{NAV_{f,t-1}} - 1 \quad (2)$$

$NAV_{f,t}$ is the net asset value of fund f at time t and D_t is the dividends paid out by the fund in period t. The return spreads for the factors are available through the Kenneth R. French Data Library[11].

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

4.2.1.1 Alternative performance measures

This section continues with a robustness check, where the CAPM, FF3, and CH4 are used to compare the results of FF5. CAPM is a model that connects the required rate of return on a security to its systematic risk measured by beta (Brealey et al., 2009). The CAPM model captures the portfolio's excess return over what is expected from them based on its systematic risk (Barnett & Salomon, 2006; Capelle-Blankcard & Monjon, 2012; Renneboog et al., 2008b). Jensen's alpha (1968) is added to the CAPM model to predict the portfolio's return above or below what the CAPM model expects. According to the CAPM model:

$$r_{i,t} - r_{f,i} = \alpha_{J,i} + \beta_{m,i}\left(r_t^m - r_t^f\right) + \varepsilon_{i,t} \quad (3)$$

Where $r_{i,t}$ signifies the return of fund i in month t. $\alpha_{J,i}$ is the Jensen's alpha (Jensen, 1968); this alpha resembles the interception of the return beyond what is predicted by the corresponding model. r_t^m symbolises the return on the market portfolio for month t, r_t^f is the risk-free rate of return on a one-month United States T-bill. $\beta_{m,i}$ captures the fixed beta of fund i. $\varepsilon_{i,t}$ is the random error.

Additionally, the FF3 and the CH4 are used to calculate the RAP. The FF3 captures the relationship between average return and size and return and price rations (Fama & French, 1993). The CH4 model adds one factor to the FF3 formula: the momentum factor for asset pricing. The calculations for FF3 and CH4 are:

$$r_{i,t} - r_{f,i} = \alpha_{3f,i} + \beta_{m,i}\left(r_t^m - r_t^f\right) + \beta_{smb,i} * r_t^{smb} + \beta_{hml,i} * r_t^{hml} + \varepsilon_{i,t} \quad (4)$$

$$r_{i,t} - r_{f,i} = \alpha_{4f,i} + \beta_{m,i}\left(r_t^m - r_t^f\right) + \beta_{smb,i} * r_t^{smb} + \beta_{hml,i} * r_t^{hml} + \beta_{mom,i} * r_t^{mom} + \varepsilon_{i,t} \quad (5)$$

Where $\alpha_{3f,i}$ and $\alpha_{4f,i}$ are the regression interceptions and will be referred to as the three-factor and four-factor alpha. r_t^{smb} is the return of the small-minus-big (SMB) factor that is proxied by the return spread between a small-cap portfolio and a large gap portfolio. Equivalent, r_t^{hml} is the return of the high-minus-low (HML) factor, the gap between a high B/M ratio and a low B/M ratio portfolio. Furthermore, r_t^{mom} is the momentum factor (MOM), the difference between a portfolio holding winners and a portfolio holding following losers.

4.2.2 Screening categories and Screening intensity

The fundamental goal of this study is to discover if screening categories and screening intensity affect the financial performance of an SRI fund. Looking at other studies, Barnett and Salomon (2006) apply five dummy variables for the screening criteria: Environment, Labour Relations, Equal Employment, Community Investment, and Community Relations. While Renneboog et al. (2008b) combined information from various data sources to identify 21 screening criteria, they classified them into four major categories: Sin, Ethical, Corporate Governance and Social, and Environment.

The independent variables for this study are the screening intensity and a set of screening categories. The screening intensity is determined by the amount of (sub-)screening criteria the fund uses to screen the investments. This intensity is measured as the absolute number of screens used (Barnett & Salomon, 2006; Renneboog et al., 2008b; Lee et al., 2010). The total number of (sub-) criteria given by USSIF that an SRI mutual fund can apply is 17; this number will also be the maximum intensity a fund can utilise. If a fund has a screening intensity of 17, it implies that it uses all the 17 criteria and will screen for all the primary categories.

The USSIF website provides six screening categories for investors to select mutual funds. These six categories are Environment, Social, Governance, Products, Other/Qualitative, and Shareholder engagement. The website divides these overarching categories into sub-criteria; table 1 below shows the six primary categories and their criteria. For a fund to screen for a primary category, it needs to screen for all the sub-criteria. All the screening categories are added as dummy variables to the regression, where one means a fund screen for that particular category. For example, a fund that screens for the Environmental category apply all three screening criteria: climate/cleantech, pollution/toxics, and environment/other. If this is not the case, then a zero will be documented for the Environmental dummy.

Table 1: Screening categories used for this study.
This table presents the screening categories Environment, Social, Governance, Products, Other/Qualitative, and Shareholder engagement with its screening criteria. https://charts.ussif.org/mfpc/ provides the screening categories with their corresponding screening criteria.

Screening category	Screening criteria
Environment	Climate / Clean tech
	Pollution / Toxics
	Environment / Other
Social	Community development
	Diversity & EEO
	Human rights
	Labour relations
	Conflict risk
Governance	Board Issues
	Executive Pay
Products	Alcohol
	Animal welfare
	Defence / Weapon
	Gambling
	Tobacco
Other/ Qualitative	-
Shareholder Engagement	-

4.2.3 Control Variable

To gain a thorough understanding of the hypothesised relationships, the model needs to include control variables that could systematically affect the financial returns of SRI mutual funds. Following the stream of literature examining the relationship between screening criteria and the section devoted to general mutual funds, this study includes the fund's characteristic control variables age, size, expense ratio, and global fund (Barnett & Salomon, 2006; Capelle-Blancard and Monjon, 2012; Lee et al., 2010 Renneboog et al., 2008b).

The first control variable will be the ***age*** of the fund. Existing literature provides evidence that the age of a fund negatively impacts the RAP of funds (Renneboog et al., 2008b; Singh & Tandon, 2021; Webster, 2002). Age is included in the model as the natural logarithm year-end. (Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Humphrey and Lee, 2011; Lee et al., 2010; Renneboog et al., 2008b; Renneboog et al., 2011). The logarithm is used to overcome outliers in the regression analysis.

The second variable is ***size***. As mentioned above, Keswani et al. (2013) found that the size of a fund is negatively related to its performance in the United States. Therefore, the variable size will include the natural logarithm of total net assets under management reported at the end of each year to correct the size effect (Barnett & Salomon, 2006; Capelle-Blancard and Monjon, 2012; Renneboog et al., 2008b; Renneboog et al., 2011).

The ***expense ratio*** of funds negatively affects SRI funds' risk-adjusted performance (Carhart, 1997; Renneboog et al., 2008b). The variable is the total annual costs associated with investing in the fund expressed as a percentage of the investment. The variable is lagged since it could predict the next period's return (Pástor et al., 2015).

Lastly, the macro-economic effects of investing internationally versus only investing in the domestic market can affect the financial returns (Barnett and Salomon, 2006). The model includes a dummy variable ***global fund*** for funds with international holding and zero otherwise to control macro-economic effects (Barnett & Salomon, 2006).

4.3 Statistical methods

The main objective is to study the effect of individual screening criteria on the financial performance of SRI funds. This research follows a methodology similar to existing literature to facilitate comparison. Appendix III shows various literature on this topic with the method used. The primary screening types are Environmental, Social, Governance, Products, Other/Qualitative, and Shareholder Engagement. The use of screens can be regarded as an active selection strategy to generate superior fund performance. Therefore, the number and type of SRI screens in our model are included to explain SRI funds' RAP. Additionally, the variables above can explain the fund's performance; hence, our model of SRI fund returns looks as follows:

$$RAP_{i,t} = \gamma_0 + \gamma_1\, Screening\ Category_{k,i} + \gamma_2 Fund\ Characteristics_{i,t-1} + \gamma_3\, Screening\ Intensity_i + \gamma_4 (Screening\ Intensity_i) + \varepsilon_{i,t} \quad (4)$$

RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of 12 months, adjusted by specific risk factors. FF5 model is used to calculate the RAP. $Screening\ Intensity_i$ is equal to the number of criteria used by fund i. The model will also contain the square of the number of criteria used to capture a potential nonlinear relationship (Capelle-Blancard and Monjon, 2012). $Screening\ Category_i$ is a set of six dummy variables: $Environmental_i$, $Social_i$, $Governance_i$, $Products_i$, $Other/Qualitative_i$, and $Shareholder\ Engagement_i$. The dummy variable is one if fund i utilizes all criteria of category k, where k is the individual categories used as described in the former chapter.

$Fund\ Characteristic_{i,t-1}$ are variables consistent with the control variables mentioned above. Where (1) $Size_{i,t-1}$ is the natural logarithm of fund assets under management in Dollars at year t-1; (2) $Age_{i,t-1}$, is the natural logarithm of the number of years since the fund's data of establishment; (3) $Global\ fund_{i,t-1}$ is one the whether the fund trades international, or 0 when the fund does not trade international; (4) $Expense\ ratios_{i,t-1}$, the lagged annual expense ratio in percentage. (5) $\varepsilon_{i,t}$ might not be independent across time due to the macro-economic factors associated with performance (Barnet and Salomon, 2006).

5. Sample composition and data collection

This chapter provides information on how the sample is constructed and gathered. The fund-level data of United States SRI mutual funds are collected from the USSIF website and Refinitiv Eikon database. The factors used to calculate the RAP are gathered from Kenneth R. French data library, which provides benchmark returns for the last month, quarter, and year and downloadable files with historical returns going back to 1926[12].

This study is not the first to use the USSIF platform; both Donath et al. (2018) and Lee et al. (2010) have used the USSIF platform to research SRI funds. The data from USSIF is publicly available for institutional members and individual investors who wish to find financial performance, costs, voting records, and screens of all the listed SRI funds. USSIF also provides data on the fund age, total AUM, the expense ratio of the SRI funds, and whether the fund is trading internationally. However, the platform only supplies current data and does not provide historical information. To clarify, the information obtained from USSIF are the names of the SRI funds, the screening criteria applied, and the intensity of screening. All the other information needed for this research is gathered from the Refinitiv Eikon database. The first section of this chapter provides information on the sample's composition. After which, the collection of data from Refinitiv Eikon is discussed.

5.1 Sample composition

The USSIF platform provides a list of 180 sustainable investment mutual funds and ETFs in the United States. Bloomberg LP provides the financial performance data on the website. Firstly, only equity mutual funds will be used in the dataset for this research. This selection ensures better comparability, as mutual funds that invest in bonds and other financial instruments have different investment strategies (Eltong, Gruber, & Blake, 1996). However, funds are included in the sample that invests a small amount in bonds. After filtering, 68 SRI funds remain. Some funds offer different share classes; the selection will only have each A-class share of each fund. The removal of duplicates improves the comparability; for example, the sample includes Calvert Equity Portfolio A, while the selection excludes Calvert Equity Portfolio C and I, the asset under management (AUM) of all three shares is identical.

Additionally, the sample excludes all the institutional shares since the objective of this research concentrates on the shares retail investors can include in their portfolios that align their beliefs and provide a financial return. All the closed-end fund shares are omitted from the selection because they trade in the open market, and their actual prices fluctuate according to demand and supply; these shares can sell at a discount or premium of NAV. Lastly, funds with incomplete data from 2015-to 2020 are omitted from the sample. The final selection contains 41 SRI funds. Deleting missing observations can result in biased parameters and estimates and reduce the analysis's statistical power. The missing data seem random; this might be because of the database failure or because the samples are lost in transit or technically unsatisfactory. Therefore, the biased created by omitting variables might be very small or non-existing (Hair, Black, Babin & Anderson, 2014).

[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

5.2 Data

The data is collected using the USSIF website and the Refinitiv Eikon database. USSIF is a publicly available website, while Refinitiv Eikon provides real-time and historical market data and is accessible through a license from the University of Twente.

The collected data for this research covers Jan. 2015 till Dec. 2020. There are a couple of reasons for this period. Firstly, it covers a period without the global financial crisis of 2008 and its aftermath. Second, choosing a recent period contributes to the applicability of this study. Third, there is no empirical evidence specifically for this period to the best of the author's knowledge. Selecting a recent period contributes to the applicability of this study and adds to the existing literature that uses a period of at least five years (Capelle-Blancard & Monjon, 2012; Donath, Ioan & Mandimutsira, 2018).

As discussed, this study only selects funds with data available for all variables of interest in all consecutive years. Therefore, the sample data is a longitudinal and balanced panel data set. Since this study uses balanced panel data, each panel member (N) is observed annual (T). The total amount of observations (n) is equal to $n=N*T$. After applying all the criteria, the final dataset collected from Refinitiv Eikon resulted in 41 SRI funds. The total amount of observations for this study is:

$246\ (n) = 41(N) * 6\ (T)$[13]

5.3 Screens

The USSIF website contains all the individual screens used for this research. USSIF and Bloomberg define screens used in the investment process for each SRI fund. The screening data will correspond with the data retrieved from USSIF. Unfortunately, USSIF does not provide historical data. Therefore, the study might suffer from look-ahead bias since the information is paired with historical data from the other database. Later on, this paper will discuss the implication of the look-ahead bias in more detail. Moreover, it is essential to mention that this paper assumes that the screening mandate is difficult to change over time since USSIF does not provide historical data on the screening criteria of SRI funds. Table 1 in section 4.2.2 above provides the information on the screening criteria gathered from USSIF with the different sub-divisions used to determine if the fund screen for a particular criterion.

[13] The minimum ratio of obersvations to variables is 5:1, but the preferred ratio is 15:1 or 20:1 (Hair et al., 2014). With 11 independent and control variables our ratio is around the 22:1, which satisfies the sample size considerations.

6. Results

The following part presents the results of the empirical analyses. This chapter starts with a statistical summary of the variables. After which, the paper provides the specification of all the assumptions and conditions of regression analysis. The correlation analysis is discussed in a different section to look at the collinearity between the variables. The regression results will follow this section and the final part contains the robustness tests to increase the overall research quality.

6.1 Univariate analysis

Table 2 (p. 27) below shows the United States socially responsible investment funds sample's descriptive statistics from January 2015 to December 2020. The table includes the descriptive statistics of the dependent, independent, and control variables used in the regression analyses. The different calculations for the RAP variable are also added to the table to compare the results. These summary statistics describe the essential characteristics of the data. The entire sample consists of 41 SRI mutual funds with 246 observations over six years. To the best of the author's knowledge, no SRI fund ceased operations during the sample period. Therefore, there is no need to correct survivor bias in the data (Capelle-Blancard & Monjon, 2014).

While performing the univariate analysis, outliers were identified and addressed to resolve any issues. Outliers are observations with unusually high or low values on a variable or a unique combination of values across several variables that make the observation stand out (Hair et al., 2014). Variations in the return could easily but falsely contribute to marginal changes in the results. The framework of Adems, Hayunga, Mansi, Reeb, and Verardi (2019) is applied to deal with the outliers. The variable RAP is winsorised at the 5% level to eliminate all outliers. The winsorising value of 5% is used because the 1% level did not eliminate all the extreme outliers[14]. The variable expense ratio is not winsorised since the analysis did not identify outliers. All variables with a natural log transformation do not need to be winsorised, as natural logarithms overcome possible outliers in the regression analysis (Hair et al., 2014). The RAP variable in the descriptive statistics is winsorised; the other variables presented in Table 2 are *before* winsorising and log transformation.

The mean monthly return of the sample is 0.83%, with a minimum of -97.5% and a maximum of 123.5%. The monthly return results in an annual mean return of 10.43%[15]. The table provides insight into the various returns adjusted for the different risk factors. RAP is the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a given month based on 12 months, adjusted by specific risk factors. The mean return of CAPM, FF3, and CH4 is -0.076% monthly, corresponding to -0.908% annual[16]. This result implies that over the past 6-years, the SRI funds have had an average negative return of -0.076% monthly, adjusted for risk. By adding the two new factors of FF5, namely investment and profitability, the mean return slightly increases to -0.004% monthly (-0.048% annual[17]). Likewise, Renneboog et al. (2008b) reported negative alphas for their research. They found for the CAPM a return of -2.84% annual and for CH4 -3.37% annual between 1991 to 2003. On the contrary, Barnett and Salomon (2006) had a mean return of 0.13% monthly using the CAPM model between 1972 to 2000. Their monthly CAPM is equivalent to 1,57% annual.

Furthermore, table 1 shows that the sample fund age varies between 3 and 39 years, with a mean of 15 years and an average size of $508.87 million (min of $1.6 million and a max of $10,407.90 million). Age is measured in years from inception, and size is the total AUM. Compared with other studies, Renneboog et al. (2008b) studied 98 United States SRI funds and reported an age mean of 7 years, with an average AUM of $142.1 million. Moreover, Capelle-Blancard and Monjon (2014)

[14] Logging the RAP variable did not help with the excessive kurtosis levels that the data showed; therefore, winsorising the data was the best way to deal with the outliers.
[15] ((1+0.0083)^12)-1) * 100 = 10.43% annum
[16] ((1-0.00076)^12)-1) * 100 = -0.908% annum.
[17] ((1-0.00004)^12)-1) * 100 = -0.0479% annum.

investigated French SRI funds and had fund age vary between 3 years and 25 years, with a mean of 8 years, with assets ranging from €0.32 million to €1,446.21 million and an average of €116.35 million. Lastly, Barnett and Salomon (2006) studied U.S. SRI funds and reported an average AUM of an average of $93 million (min $0.19 and max $1483.92 in millions). Interestingly, the average size of the funds has increased compared with older literature. The increased SRI market could explain the increase in the size of SRI funds.

The expense ratio (EXP) is the following control variable and is the total annual costs associated with investing in the fund expressed as a percentage of the investment. The SRI funds in our sample have an average expense ratio of 1.066%. The lowest expense ratio applied by SRI funds in our selection is 0.32% (min), while the highest expense ratio is 1.85% (max). The expense ratio is the cost of owning a mutual fund, meaning that, on average, an SRI investor pays a fee of 1.066% on their investment in the SRI mutual fund. Renneboog et al. (2008b) show an average total expense fee of 1.9% of United States SRI funds from 1991 to 2003. Although Capelle-Blancard and Monjon (2014) used management fee instead of expense ratio, a fund's management fee is simply a portion of a fund's overall expense ratio. The researchers recorded for management fee a mean of 1.20%. These comparisons indicate a slight decrease in the average expense ratio for investors to pay.

The control variable global (GLO) signals if a fund invests internationally. The global dummy has a mean of 0.34. Barnett and Salomon (2006) reported a mean of 0.07 for their global variable. This increase in the worldwide fund variable can imply that United States SRI funds are trading more globally than before. It is also interesting to note the relative popularity of types of screening criteria. Many SRI funds screen for the environment (0.9), and less than half of the SRI funds concern themselves with social screening (0.46). It is more challenging to compare all the screening criteria with existing literature because most studies use different screening criteria and measure them differently. Nevertheless, Barnett and Salomon (2006) also screened for the environment and stated a mean of 0.83, slightly lower than our mean of 0.90. This difference might imply an increase in environmental awareness.

Table 2: Descriptive Statistics on the variables

This table provides the descriptive statistics of the sample of 41 SRI funds ranging from January 1st, 2015, until December 31st, 2020. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end on a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is years from inception. The expense ratio is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Size is the total assets under management in U.S dollars and millions. The global dummy signals if part of the fund's assets are invested internationally. The abbreviations used in the table are minimum (Min), and maximum (Max), standard deviation (Std. Dev).

Table 2: Descriptive Statistics on the variables used

	Min	Max	Mean	Std. Dev
Dependent variable				
Monthly return in %	-0.975	1.235	0.830	0.074
Independent variables				
SI	10	15	13.68	1.618
D_Env	0	1	0.90	0.297
D_Soc	0	1	0.46	0.500
D_Gov	0	1	0.83	0.377
D_Pro	0	1	0.88	0.328
D_Qua	0	1	0.34	0.475
D_Sha	0	1	0.88	0.328
Control variables				
Size	1.600	10407.90	508.871	1496.154
EXP	0.32	1.85	1.066	0.337
AGE	3	39	15.50	7.917
GLO	0	1	0.34	0.475
RAP				
CAPM	-0.224	0.086	-0.076	0.074
FF3	-0.249	0.135	-0.076	0.076
CH4	-0.248	0.235	-0.076	0.078
FF5	-0.043	0.054	-0.004	0.022

6.3 Bivariate analysis and assumption

This part of the study provides information on all the requirements and assumptions that need to be met before performing regression analyses. Section 6.3.2 discusses multicollinearity and presents the correlation matrix in table 3. The matrix is discussed in a different part because multicollinearity is a common thread in these types of research, and therefore, it would be clearer to discuss this problem in a separate section in more detail.

6.3.1 Assumption and conditions

Multivariate techniques are all based on a fundamental set of assumptions representing the requirements of the underlying statistical theory (hair et al., 2014). Hence, the data must satisfy several assumptions and conditions before performing a multiple regression. Those assumptions are the normality, independence, equal variance, and multicollinearity assumptions (hair et al., 2014). For brevity, tables and graphs are not presented in this study.

An implicit assumption of all multivariate techniques is linearity. The linear model predicts values that fall in a straight line by having a constant unit change (slope) of the dependent variable for a constant unit change of the independent variable (hair et al., 2014). The most common way to assess linearity is to examine scatterplots of the variables and identify any data's nonlinear patterns. The scatterplots of the data do satisfy the linearity assumption.

The following assumption relates primarily to dependence relationships between variables. Homoscedasticity refers to the notion that dependent variables exhibit equal levels of variance across the range of predictor variables (hair et al., 2014). This assumption can be checked by plotting the residuals and predicted values to see if the output looks like a random array of dots. The graph in our study shows some form of pattern. Robust regression analysis is applied to determine if this violation changes the results[18].

Additionally, the residuals must be normally distributed to meet the normality assumption. This assumption can be tested by looking at the P-P plot for the model. This study's P-P plot shows a slight deviation from the diagonal line, suggesting that the assumption of normality of the residuals may have been violated. However, as only extreme deviations from the normality are likely to impact the findings significantly, the results are probably still valid (hair et al., 2014).

6.3.2 Correlation analysis

The Pearson's correlation matrix was used to test whether the relationship between the variables was statistically significant. The Pearson's correlation matrix provides information on the magnitude of the association, or correlation, and the direction of the relationship. Pearson's R is also helpful to check for multicollinearity. The coefficient values can range from +1 to -1, where +1 indicates a perfect positive relationship, -1 implies a perfect negative relationship, and 0 indicates no association exists.
This type of analysis is primarily used to locate multicollinearity between the independent variables in the regression analysis. Multicollinearity can distort the regression analysis results when there are high correlations between two explanatory variables. Table 3 below presents an overview of the correlation between the variables used in this study. The first column of Table 3 shows the relationship between RAP, the independent, and control variables. As shown in table 3, there is no significant correlation between the dependent variable RAP and the other variables.

However, some variables have a high correlation, especially the screening criteria correlated with each other. For example, Environmental and Governance screens connect with r =0.725. This correlation implies that funds use these screens in more conjunction with each other. Additionally, table 3 shows that screening intensity is highly associated with the screening criteria' Environment, Social, Governance and Shareholder Engagement. This relationship seems reasonable since an increase in the screening criteria implies that a fund had added one of the screening categories.

[18] The results did not change, the robust regression analysis can be found in section 6.5 Robustness check

Besides, the age of a fund is positively correlated to the size of a fund (r=0.542). This correlation seems logical as it implies that the older a fund gets, the size of the fund increases.

The main concern is that higher correlations may infuse multicollinearity into the regression. To deal with this problem of high correlations, this research will follow the example of Barnett and Salamon (2006) and perform various sensitivity analyses to deal with the issue of multicollinearity and ensure the robustness of the results. Firstly, each screening criteria will be added separately to the regression. Secondly, the order of independent variables will be changed and added to the analyses. And lastly, this research constructs models with screening criteria that are not highly correlated with each other to detect if the outcomes will vary.

Table 3: Pearson's R correlation matrix

This table shows the Pearson's correlation matrix, including the statistical significance level of the key variables used in this research. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio (EXP) is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is invested internationally. All correlations above .500 are bold.

Table 3: Pearson's R correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1 RAP	--											
2 SI	-0.045	--										
3 D_Env	-0.007	**0.750**	--									
4 D_Soc	-0.090	**0.667**	0.306	--								
5 D_Gov	-0.039	**0.794**	**0.725**	0.422	--							
6 D_Pro	0.041	0.204	-0.123	-0.102	-0.169	--						
7 D_Qua	-0.068	0.396	0.237	**0.569**	0.327	-0.203	--					
8 D_Sha	-0.120	**0.665**	**0.882**	0.346	**0.623**	-0.139	0.268	--				
9 Log_SIZE	-0.180	0.367	0.345	0.227	0.330	-0.046	0.461	0.344	--			
10 EXP	0.123	0.000	-0.050	-0.121	-0.001	0.400	-0.283	-0.133	-0.254	--		
11 Log_AGE	-0.193	0.052	0.091	-0.086	0.197	-0.018	0.183	0.190	**0.542**	-0.143	--	
12 GLO	0.110	-0.018	0.063	-0.050	0.190	-0.203	-0.193	-0.046	-0.117	0.449	-0.049	--

6.4 Regression analysis

This section aims to provide empirical evidence for this research by testing the hypotheses formulated in chapter 3. The first subsection offers the results regarding the impact of screening intensity on the risk-adjusted return of SRI mutual funds (hypothesis 1). The second part presents the results regarding the effects of different screening categories on the performance of an SRI fund (hypotheses 2,3,4, and 5). Tables 4 to 6 show the results of the OLS regression analyses. While interpreting these results, please keep in mind the implicit assumption that SRI funds' screening procedures remain constant during the entire period.

6.4.1 Results on Screening Intensity

The first focus of this study is on the relationship between the screening intensity of SRI funds and their effect on financial performance. The first hypothesis stated that the relationship between SRI funds' intensive screening and financial performance is not u-curved. This hypothesis is that SRI funds can diversify their portfolios better because ESG has become more mainstream. Figure 1 below presents the first look at the relationship between the monthly returns and the number of screening criteria. The simple bivariate analysis shows that screening intensity and monthly return do not have a linear relationship. At a screening intensity of 10, the SRI firms made a positive return of roughly 0.5%; however, applying one more screen, the return drops to a negative 2% at 11 screens. After which, the return increases again at 12 and 13 screening criteria. A screening intensity of 14 offers the highest monthly return of almost 4% monthly, corresponding to an annualised return of 60.1% over six years[19]. After 14 screens, the return decreases when a firm adds more screens; at 15 applied screens, the return is back at approximately 0.5%. At 16 screens, the return increase slightly, at which it almost stabilises between 16 and 17 screenings.

These results might indicate that screening intensity and financial performance are curvilinear. Remarkably, the figure does not show a U-curve as found in the literature (Barnett & Salomon, 2006; Capelle-Blancard & Monjon, 2014) but an inverse n-shaped line. However, there need to be more sophisticated analyses to provide proof for these first findings.

Figure 1: Simple Line Mean monthly return by SI

The table displays a simple line with the Mean monthly return in percentage and the screening intensity. Screening intensity is measured as the absolute number of screens used by each fund. Where screening intensity ranges from 1 to 17, the dataset includes 41 U.S SRI mutual funds from January 2015 until December 2020.

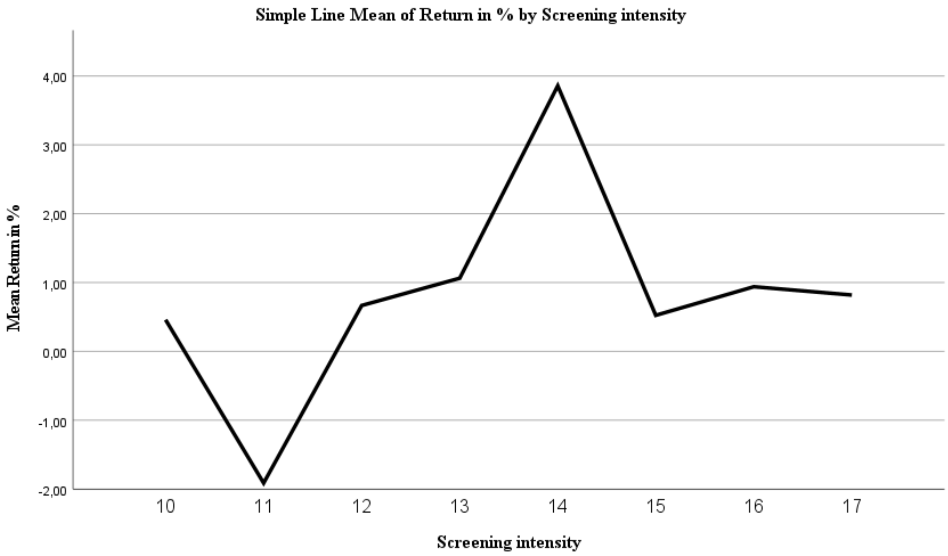

Table 4 below shows the OLS regression models testing the first hypothesis about screening intensity. Model 1 posits the RAP as a linear function of screening intensity. Specifically, it tests whether including more social screens is positively or negatively related to financial performance. A negative relationship would support the critics against social screening, while a positive relationship would support the opponents of socially responsible fund's outperformance. Model 2 adds a squared screening intensity to indicate a non-linear relationship. Model 1 and 2 are nested and therefore can be directly compared. The adjusted R square increased from model 1 to model 2 and is significant at the 5% level in both cases. Screening intensity is harmful and not significant in model 1 and changes to a positive coefficient; however still insignificant. The squared screening intensity added in model 2, with a coefficient of almost 0, is also not significant.

Interestingly, the results contradict the findings of Barnett and Salomon (2006) and Lee et al. (2010). Barnett and Salomon (2006) concluded a negative and significant coefficient for screening intensity and a positive significant coefficient for the quadratic, meaning that the relationship between the number of screens and the financial performance is u-curve. The results are not significant, and therefore, there cannot be concluded that the line shows an n-curve or a u-curve.

[19] ((1+0.04)^12)-1) * 100 = 60.1% annum.

In all of our tests, the square of screening intensity is not substantial. Therefore, this research could not find evidence of a curvilinear relationship between screening intensity and the RAP of United States SRI funds. The findings support the first hypothesis that the relationship between screening intensity and risk-adjusted performance is not curvilinear. The results align with research done by Laurel (2011), Lee et al. (2010), and Renneboog et al. (2008b). They also concluded no significant relation between screening intensity and financial performance. Hence, retail investors do not have to concern themselves with the screening intensity of SRI mutual funds as this does not add to or decrease the risk-adjusted performance.

Table 4: Regression results for screening intensity

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the OLS regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio is the total of annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Table 4: Regression results for screening intensity

	Model 1 (RAP)	Model 2 (RAP)
Constant	0.013	-0.049
	(0.336)	(0.404)
SI	-1E-04	0.009
	(0.900)	(0.286)
SI2		0.000
		(0.279)
Log_SIZE	-0.001	-0.001
	(0.346)	(0.382)
EXP	0.003	0.004
	(0.486)	(0.438)
Log_AGE	-0.006*	-0.006*
	(0.071)	(0.067)
GLO	0.003	0.003
	(0.315)	(0.408)
Adjusted R2	0.036	0.037
F-statistic (overall)	2.823**	2.746**

* *Statistical significance at the 10% level*

** *Statistical significance at the 5% level*

*** *Statistical significance at the 1% level*

6.4.2 Results on Screening Category

This section discusses the hypotheses' results concerning the different screening categories SRI mutual funds apply. Table 5 to 7 presents the screening category results and the effect on financial performance. The results show that some screening categories significantly influence mutual funds' performance. Remarkably, some of the relationships are not as hypothesised. Table 5 presents the individual screening criteria with the dependent and control variables. The last model of table 5 portrays the results when all the screening criteria are in one model. Furthermore, table 6 adds each screening criteria step-by-step, and finally, table 7 shows the screening criteria paired based on noncorrelation created by table 3, the correlation matrix, to detect if multicollinearity changes the results.

The second hypothesis articulated in this research relates to the Environmental screening category. It formulates that screening based on environmental screens would not earn additional financial returns. Indeed model 1 of Table 5 displays a nonsignificant relationship between the RAP and environmental screening criteria without the other measures added to the model. However, model 7 of Table 5 shows that the environmental screening is significant at the 1% level, with a 0.033 coefficient, implying that this screen could contribute 0.033% to the RAP monthly. This percentage corresponds to a 0.397% annual increase in the environmental screen on the RAP[20]. These results contradict the second hypothesis and provide evidence that the environmental screening criteria positively influence the risk-adjusted performance of SRI funds in the United States. This finding contradicts the results from existing literature. For instance, Barnett and Salomon (2006) concluded that environmental screens negatively impact the risk-adjusted performance of SRI funds. Moreover, Auer (2014) found that environmental screens do not significantly add value.

The social screening category is the third theory formulated in this research. The hypothesis states that screening based on social criteria would earn SRI funds a higher financial return. The social screening criteria seem to lack statistical significance. In models 2 and 7, the coefficients are both -.004%. The findings do not support the hypothesis, and with it, the conclusion is drawn that the benefits of good employment relationships seem to be offset by the costs incurred by these firms. These results are aligned with research done by Auer (2014). The researcher found that social screening criteria do not achieve additional RAP.

The fourth hypothesis concerned the governance category, which stated that the governance criteria would earn SRI funds a higher financial return. Contrary to this belief, table 5 shows that the governance variable is nonsignificant. The findings do not support hypothesis 4, concluding that, based on this research, the governance requirements do not add a RAP to SRI funds in the U.S than funds that do not screen for the governance criteria. This result contradicts the findings from Auer (2014), who concluded that the governance screen was the most important criterion for achieving higher performance.

Table 5 shows that the product category also lacks statistical significance. These results align with hypothesis five, which stated that product criteria would not earn SRI funds a higher financial return than funds that do not screen for products. The nonsignificant relationship might be because fund managers can expose their portfolios to the same factors as sin stock to regain their expected return (Blitz & Fabozzi, 2017).

The results on shareholder engagement criteria are slightly different to what was theorised. Hypothesis 6 stated that this criterion would not earn SRI funds a higher financial return. The results show even a statistically significant negative relationship at the 1% level in model 7. Shareholder engagement has a significant coefficient of -0.027%. This coefficient means that funds participating in shareholder engagement suffer a monthly financial penalty of 0.027%, corresponding to -0.324% annually[21]. The results show that managers of SRI funds who engage in dialogue, resolutions and proxy voting negatively affect the RAP of their fund. These results are supported by Haigh and

[20] ((1+0.00033)^12)-1) * 100 = 0.3967 percent per annum.
[21] ((1-0.00027)^12)-1) * 100 = -0.324 percent per annum.

Hazelton (2004), who argue that SRI funds lack the power to create significant corporate changes and that shareholder advocacy has been largely unsuccessful. Therefore, retail investors should avoid SRI funds that engage in these efforts

Noticeably, Environmental and Shareholder Engagement are only significant in model 7 when all the criteria are placed in the same model. Table 3 shows that the two variables are highly correlated, with a Pearson's R of 0.882. Therefore, the next step is to perform different analyses to provide more insight into the results and if multicollinearity may change the results.

Table 5: Results of screening criteria

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 United States mutual funds. This table presents the OLS regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy signals if part of the fund's assets are invested internationally.

Table 5: Individual results on the screening categories

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)	Model 7 (RAP)
(Constant)	0.010	0.015	0.012	0.011	0.012	-0.014	0.007
	(0.325)	(0.117)	(0.203)	(0.229)	(0.205)	(0.129)	(0.488)
D_Env	0.003						0.033***
	(0.566)						(0.007)
D_Soc		-0.004					-0.004
		(0.219)					(0.251)
D_Gov			0.000				-0.003
			(0.956)				(0.642)
D_Pro				0.003			0.003
				(0.575)			(0.540)
D_Qua					0.002		0.006
					(0.655)		(0.174)
D_Sha						-0.004	-0.027***
						(0.354)	(0.006)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002*
	(0.210)	(0.489)	(0.279)	(0.252)	(0.228)	(0.410)	(0.087)
EXP	0.003	0.003	0.003	0.001	0.004	0.003	0.001
	(0.486)	(0.545)	(0.489)	(0.800)	(0.455)	(0.521)	(0.914)
Log_AGE	-0.005*	-0.007**	-0.006*	-0.006*	-0.006*	-0.006*	-0.004
	(0.085)	(0.039)	(0.071)	(0.068)	(0.078)	(0.071)	(0.276)
GLO	0.003	0.003	0.003	0.004	0.003	0.003	0.003
	(0.346)	(0.305)	(0.331)	(0.274)	(0.299)	(0.305)	(0.369)
Adjusted R2	0.037	0.042	0.036	0.037	0.039	0.059	0.065
F- statistic (overall)	2.890**	3.142***	2.820**	2.888**	2.102**	4.056***	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

The first sensitivity measure is to add each screening criteria step-by-step to the regression, presented in table 6. This table shows that the environmental screen only becomes significant at the 1% level if the Shareholder screen is added to the regression [22]. As mentioned above, Environmental screening and Shareholder engagement are correlated with each other. These results imply that investors should look for SRI funds that screen for the Environment and actively engage as shareholders. Even though the shareholder engagement negatively affects the risk-adjusted return, the Environmental screen adds more to the financial performance than the Shareholder engagement reduces. A positive coefficient of 0.033% and a negative coefficient of -0.027% would improve the RAP by 0.006% monthly, equivalent to 0.072% annual[23]. In model 5, the Social screen shows a small significant effect on the RAP at a 10% level. The variable is insignificant in model 1 till 4; when the model excludes the Shareholder Engagement in model 5, the social screen becomes slightly significant. Again, the variable is insignificant in the last model 7, including all the variables.

Based on Barnett and Salomon (2006), another sensitivity analysis to investigate if changing the order of the independent variables changes the significance of the variables on the RAP of SRI funds. Appendix V shows the same type of tables as tables 5 to 7 in this section. However, in the appendix, the order of the independent variables is changed. In appendix V, Shareholder Engagement is the first to enter the regression, and Environment is the last. The results do not differ from the main results; the order of the independent variables does not impact the outcome.

The following sensitivity analysis is constructed based on the Pearson's R correlation matrix displayed in table 3. The requirement for including the independent variables in the models was that the Pearson's R correlation was not more extensive than R = 0.5[24]. Table 7 shows the results of various models that exclude highly correlated independent variables.

Model 1 of Table 7 provides developments on Social screening; when this criterion is paired with Environmental, Product, and Qualitative screening categories, the variable provides a significant negative impact on the RAP. However, the significance level is weak at a 10% level, with a *p* value of 0.053. The social screen negatively affects the RAP with -0.007% monthly, equivalent to -0.084% annual. However, the screening variable loses its significance in model 2, where the variables used in the model are changed. Interestingly, Tables 6 and 7 show that the social screen is slightly significant when Shareholder engagement is not included in the model. In Table 6, model 5 Social screen is substantial at 10% with a coefficient of -0.007%. In table 7, model 1, a coefficient of also -0.007% at a 10% level. Both models exclude the Shareholder engagement category. Contrary, the sensitivity table 6 in appendix V shows, in model 5, that when the Environmental screen is excluded, the Social screen is also somewhat significant at a 10% level. These findings imply that the Social screen negatively affects SRI funds' performance in some combinations. When an SRI fund screens for all the criteria except for Environment or Shareholder engagement, the Social screening negatively affects the performance. However, when more variables are excluded, the impact of Social screening does not always significantly affect a fund's performance.

To conclude, the impact of the Social category changes depending on the other independent variables. However, this research cannot provide information on which, combined with the Social category, negatively impacts the RAP, and a 10% significance level is not enough to state that it significantly influences the RAP of SRI funds in the United States. This research can conclude that the Environmental screen combined with Shareholder Engagement creates value for the retail investor.

[22] OLS regressions were performed were some of the categories were removed, the results stayed the same; both D_Env and D_Sha stayed significant.

[23] ((1+0.00006)^12)-1) * 100 = 0.0720 percent per annum.

[24] The results did not change when checked if the correlation between the variables AGE and LOG_Size would impact the results.

Table 6: Screening category step-by-step

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The sample consists of 41 SRI mutual funds located in the United States. This table presents the OLS regression results on the relationship between the dependent variable risk-adjusted performance and the independent variable screening category. RAP, calculated with FF5, and the separate variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy signals if part of the fund's assets are invested internationally. All returns are monthly and in percentage.

Table 6: Screening criteria step-by-step

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)
(Constant)	0.010	0.012	0.012	0.011	0.011	0.007
	(0.324)	(0.234)	(0.244)	(0.291)	(0.280)	(0.488)
D_Env	0.003	0.005	0.005	0.005	0.006	0.033***
	(0.566)	(0.378)	(0.523)	(0.516)	(0.438)	(0.007)
D_Soc		-0.004	-0.004	-0.004	-0.007*	-0.004
		(0.162)	(0.191)	(0.197)	(0.066)	(0.251)
D_Gov			0.000	0.000	0.000	-0.003
			(0.989)	(0.959)	(0.946)	(0.642)
D_Pro				0.003	0.005	0.003
				(0.575)	(0.405)	(0.540)
D_Qua					0.006	0.006
					(0.143)	(0.174)
D_Sha						-0.027***
						(0.006)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002*
	(0.210)	(0.359)	(0.360)	(0.338)	(0.172)	(0.087)
EXP	0.003	0.003	0.003	0.001	0.001	0.001
	(0.486)	(0.549)	(0.550)	(0.858)	(0.862)	(0.914)
Log_AGE	-0.005*	-0.006**	-0.006*	-0.007*	-0.007**	-0.004
	(0.085)	(0.045)	(0.052)	(0.051)	(0.045)	(0.276)
GLO	0.003	0.003	0.003	0.004	0.005	0.003
	(0.346)	(0.355)	(0.371)	(0.295)	(0.189)	(0.369)
Adjusted R2	0.037	0.041	0.037	0.034	0.039	0.065
F- statistic (overall)	2.890**	2.746**	2.344**	2.084**	2.102**	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 7: Screening category paired based on noncorrelation

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The sample consists of 41 SRI mutual funds located in the United States. This table presents the OLS regression results on the relationship between the dependent variable risk-adjusted performance and the independent variable screening category. RAP, calculated with FF5, and the separate variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy signals if part of the fund's assets are invested internationally. All returns are monthly and in percentage.

Table 7: Screening criteria paired based on noncorrelation

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)
(Constant)	0.011	0.016	0.011
	(0.264)	(0.107)	(0.261)
D_Env	0.005		
	(0.321		
D_Soc	-0.007*	-0.004	
	(0.053)	(0.204)	
D_Gov		0.006	0.000
		(0.262)	(0.984)
D_Pro	0.005	0.003	0.004
	(0.401)	(0.630)	(0.504
D_Qua	0.006		0.002
	(0.142)		-0.575
D_Sha		-0.006	
		(0.293)	
Log_SIZE	-0.002	-0.001	-0.001
	(0.172)	(0.478)	(0.211)
EXP	0.001	0.001	0.001
	(0.864)	(0.857)	(0.799)
Log_AGE	-0.007**	-0.007**	-0.006*
	(0.037)	(0.034)	(0.079)
GLO	0.005	0.003	0.005
	(0.184)	(0.407)	(0.224)
Adjusted R2	0.043	0.037	0.03
F- statistic (overall)	2.374**	2.176**	2.096**

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

6.5 Robustness Check

This paper adds several robustness tests to increase the validity and reliability of the OLS regression results. The first test performs robust regression analyses to determine if heteroskedasticity affects the results; the results are presented in Appendix VI, tables 8 to 11 (p. 54-57). Another robustness check is to calculate the RAP using various performance measures; Appendix VII (p. 58-60) presents the OLS regression with the RAP calculated using the CAPM method. Appendix VIII (p. 61-63) shows the results of FF3. And lastly, Appendix IX (p. 64-66) presents the results using the CH4.

6.5.1 Robust regression

As mentioned in section 6.3.1: Assumptions and Conditions, the data does not meet the homoscedasticity assumption; this is also called heteroskedasticity. Heteroskedasticity is often seen in finance because the volatility in the price of stocks cannot be predicted over any period (Lu & White, 2014). Hayes and Cai (2007) provided an overview of heteroskedasticity-consistent (HC) standard error estimators for the least-squares linear regression model. The researcher argues that an HC estimator, preferable HC3 or HC4, should be routinely used in linear regression models. This paper will follow the advice given by Hayes and Cai (2007) and check if heteroskedasticity compromises the results. The robust models in Appendix VI, tables 8 to 11, use the HC3 estimator.

The first re-evaluated regression is that of screening intensity. Appendix VI presents the results of this re-evaluation with robust standard errors. The main results of the OLS regression for hypothesis 1 showed that screening intensity and the squared-screening intensity have an insignificant influence on the risk-adjusted performance of SRI. The robustness test results show almost identical results; both SI and SI2 are insignificant.

There is also a minor change between the OLS regression and the robust results with robust standard errors concerning hypotheses about screening criteria. Compared with the main results, variables D_Env, D_Sha and Log_SIZE all lost one significance level. In the robust analysis, the variable size does not significantly affect the RAP of SRI funds. These minor changes in the results could indicate some presence of the individual unobservable impact.

Appendix VI table 10 presents the results of the robust check on the screening criteria that are added separately to the analysis. The results show a decrease in significance for the variables D_Env, D_Sha, and Log_AGE. Model 7 indicates that D_Env and D_Sha are still significant at a 5% level, and D_Env is only meaningful in the last model. The results imply that environmental screening positively affects the risk-adjusted return of United States SRI funds. On the contrary, shareholder engagement negatively affects SRI funds' return. Lastly, Table 11 in Appendix VI presents the robust regression with the different screening groups based on collinearity. The results do not differ from the main results.

6.5.2 Alternative performance measures.

After finding that the SRI screening category can significantly impact the financial performance in the case of the Environmental screen and Shareholder engagement, this study adds a variety of supplementary calculations of the dependent variable to provide robust results. The three alternative measures used to calculate the performance are the CAPM, FF3, and CH4.

The results of the CAPM measurement are presented in tables 12 to 14 in Appendix VII. The main findings on the relationship between the RAP and screening intensity align with the outcome of the FF5 model, namely that there is no significant relationship between the fund's performance and the amount of screening they apply. Table 13 in Appendix VII provides results on the impact of the screening category on the RAP. These results show differences between using the FF5 model and the CAPM measurement. The significant Environmental screen and Shareholder engagement at the 1% level in the FF5 analysis are insignificant when the CAPM model measures performance. Table 14 in Appendix VII presents the results of the grouped screening category based on noncorrelation. The results slightly differ from the main results. The CAPM measurement shows no meaningful relationship between RAP and screening category when grouped variables; the Social screen was somewhat significant in the main results. The significant age level has dropped in all CAPM tables, wherein the FF5 model age was somewhat meaningful at a 10% level; in the CAPM model, the variable is insignificant. To finalise, after controlling for only market risk, the impact of both screening intensity and screening categories on the return is statistically insignificant

The following measurement for performance is the FF3; Appendix VIII table 15 to 17 presents the findings on the FF3 OLS regression. The findings of FF3 align with the CAPM results. The results show no statistically significant impact on the performance of SRI funds. Moreover, the FF3 models show a negative adjusted R square, indicating little explanation for the response. Therefore, controlling for size, book-to-market and excess return on the market, the impact of screening intensity and screening categories on the return stay statistically insignificant.

Both models of CAPM and FF3 show negative adjusted R squares, implying a very low or negligible explanation for the response. Negative adjusted R squared means insignificance of explanatory variables. This research cannot draw clear conclusions using these two measurements because of this negative value. The adjusted R squared may improve by increasing the sample size.

Appendix IX table 15 provides the OLS regression results that use the CH4 to analyse the relationship between the RAP and the screening intensity. The table provides the same conclusion as the main findings: screening intensity does not affect an SRI fund's RAP. The results on the screening categories when using CH4, in table 19 Appendix IX, show that the screening types do not impact the performance of SRI funds. Lastly, Table 18 provides the same evidence; all the screening categories show no significant relationship with a fund's performance. To conclude, controlling for firm size, book-to-market, and market factor, the impact of screening intensity and screening categories on the return is statistically insignificant.

However, all tables in Appendix IX show a substantial impact of age and expense ratio on the RAP, using the CH4. Fund age negatively impacts the performance of SRI funds with a significance at the 1% level. The coefficient is -0.036% monthly. This effect implies that when a fund ages, it decreases performance by 0.036% monthly, corresponding to 0.431% annual[25]. Table 16 in Appendix IX reveals that the expense ratio influences the RAP with a statistical significance at the 5% level. The expense ratio impacts the return with a positive coefficient of 0.039%. This result indicates that the RAP increases by 0.039% monthly when the expense ratio rises—corresponding to 0.47% annual adjusted for risk. After controlling for firm size, book-to-market, market factor, and momentum, the impact of screening intensity and screening categories on the return is statistically insignificant; however, age and expense ratio are statistically significant after controlling for these factors.

The negative impact of age on the risk-adjusted performance of SRI funds aligns with existing literature (Renneboog et al., 2008b; Singh & Tandon, 2021; Webster, 2002). However, the findings on

[25] ((1+ -0.00036)^12)-1) *100 = -0.431% annual

the expense ratio contradict the literature. Although most researchers found that the expense ratio negatively affects mutual funds' performance (Carhart, 1997; Renneboog et al., 2008b), the results of this study align with research done by Droms and Walker (1995). The variation in results may be due to the difference in study periods, sample size, and methodologies used, including treatments for survivor bias and benchmark issues (Prather, Bertin, & Henker, 2004). These findings on the different measurements of the risk-adjusted performance add to the meta-analysis done by Revelli and Viviani (2015), who concluded that the level of performance of SRI funds depended on the methodological choices made by researchers. This study provides insight for other researchers on the different types of performance measurement and their impact on the findings.

7. Conclusion

The central purpose of this research is to provide a comprehensive overview of the quickly developing socially responsible asset management industry and move beyond comparing SRI and conventional mutual funds. To provide more knowledge to investors who want to integrate their values into investing and make a financial return. This paper offers empirical evidence on the impact of screening intensity and screening criteria on the risk-adjusted performance of mutual funds by answering the following research question: *How do the screening intensity and category affect the financial performance of SRI funds?* The question is divided into multiple sub-questions with corresponding hypotheses to answer the research question. These relationships are investigated by performing OLS regression analyses on a dataset of 41 United States SRI mutual funds between 1st janurari 2015 to 31st December 2020, resulting in 246 observations. The main results used the FF5 to calculate the performance of the funds. The CAPM, FF3, and CH4 measurements for estimating the RAP are added to the research to provide more robust results.

Both models of CAPM and FF3 show negative adjusted R squares, implying a very low or negligible explanation. Negative adjusted R squared means insignificance of explanatory variables. Therefore, this research cannot use the results from these methods; the sample size may be too small to use these calculations for the risk-adjusted performance correctly.

The first sub-question aimed to provide knowledge on the impact of screening intensity on the risk-adjusted performance of SRI funds. A basic graph showed that the line between screening intensity and financial performance was curvilinear. Interestingly, the line was n-curve, the opposite of a u-shape found in the existing literature. This shape could be because the screening intensity has grown through the years, where Barnett and Salomon (2006) mentioned a screening intensity mean of 8 and Lee et al. (2010) reported a screening intensity scale of 1 to 11, while this study reported an average of 14 screens applied. These results could add to the findings of Barnett and Salomon (2006) that there is a turning point when adding more screens becomes harmful to the performance of a U.S SRI fund. However, the OLS regression shows that screening intensity is not statistically significant. All the different performance measurements concluded the same results on the screening intensity of SRI funds.

Consequently, this research concluded that there is no meaningful relationship between the screening intensity and the risk-adjusted performance of U.S SRI funds. Therefore, hypothesis 1 is supported, stating that the relationship between SRI funds' intensive screening and financial performance is not curvilinear. The results suggest that there are neither costs nor rewards to be gained when considering the number of screens used. Investors who want to invest in ethical mutual funds do not have to concern themselves with the amount of screening a mutual fund applies because it does not affect its return.

The following part of the research concerns the categorical screening criteria used by SRI mutual funds. The most important assumption made in this study is that the screens used by these funds remain constant over time. Using the FF5 (controlled for firm size, book-to-market, market factor, profitability, and investment patterns), *Environmental* screening criteria and *Shareholder engagement* appear to influence SRI funds' risk-adjusted return in the United States. The environmental screen positively affects the financial performance, while shareholder engagement negatively affects the performance. Various sensitivity analyses made it clear that these two variables only significantly influence financial performance when both are included in the model. Fortunately, the positive influence of the environmental screen is greater than the adverse effect of shareholder engagement. Therefore, investors should choose funds that screen for environmental criteria and active shareholder engagement to achieve a higher risk-adjusted return. The other screening criteria neither destroy nor benefit the financial performance of investors.

However, the results differ when the CH4 measurements calculate the risk-adjusted performance, controlled for firm size, book-to-market, market factor, and momentum. The results show the same effects on the screening intensity that does not impact the risk-adjusted return. Contrary to the results presented above, the CH4 does not establish any significant relationship between the screening category and the risk-adjusted performance of SRI funds. However, funds' age and expense ratio significantly impact the RAP when this calculation is used. After controlling for firm size, book-to-market, market factor, and momentum, funds age negatively influences return, and expense ratio positively affects the performance.

This study concludes that screening intensity does not impact the risk-adjusted performance. Therefore, investors do not have to worry about the number of screens mutual funds apply to investment decisions. If the Carhart four-factor model (1997) (*controlling for firm size, book-to-market, market factor, and momentum*) is used to calculate the risk-adjusted performance, the screening categories do not significantly impact the return of SRI funds. However, when the risk-adjusted is calculated using five factors (*firm size, book-to-market, market factor, profitability, and investment patterns*) with the Fama-French method (2015), the Environmental and Shareholder engagement categories significantly impact the risk-adjusted performance of SRI funds. Investors can enhance their financial return by investing in U.S SRI mutual funds that screen for Environmental criteria and are Shareholder Engagement when controlling for the five factors.

8. Limitations and future research

This last chapter begins by acknowledging the limitations of this research, followed by a discussion of the recommendations for future research regarding SRI mutual funds.

8.1 Limitations

This study is certainly no exception in having some limitations, although it tried to test the impact of screening criteria on the financial performance as thoroughly as possible. Acknowledging the drawbacks can help future studies on the risk-adjusted performance of SRI funds.

First, the results of this study are aimed at SRI mutual funds in the United States. Consequently, this does not make the results generalisable for SRI mutual funds in other parts of the world because country-specific characteristics, such as legalisation, can affect financial performance. Secondly, socially responsible investments are an upcoming topic, with a few existing pieces of literature exploring this new topic. The lack of existing literature made it difficult to set a benchmark because earlier studies found positive, negative and non-significant results on the relationship between screening criteria and financial performance. As a result, the relationship remains elusive. It was, therefore, difficult to compare our findings with those of other studies, making results more challenging to interpret.

The data retrieved from the United States Forum of Sustainable and Responsible Investment is not historical. Yet, it was cross-linked with historical time series, which indicates a look-ahead bias. However, it should be nuanced as fund mandates are difficult to change. Geczy et al. (2005) discovered that five funds from the Sustainable Investment Forum list altered their screening practices; unfortunately, they did not report the changes or the funds. These alterations in screening practices might lead to different results. Therefore, this look-ahead bias might be the most significant shortcoming in this paper and provides plenty of room for future research.

Moreover, the sample size may have been too small to use the CAPM and FF3 to calculate the risk-adjusted performance. The negative adjusted R squared means insignificance of explanatory variables, and the models could not be used for this research.

All the limitation mentioned above restricts the ability to draw a solid generalisable conclusion from this research. However, this paper still contributes to the socially responsible investment debate and sheds light on the impact of screening category and intensity on the risk-adjusted performance of SRI funds.

8.2 Further research

The topic of socially responsible investment leaves enough room for further research. As this research focused on open-end mutual funds, future research could focus on the closed-end funds that trade on the stock market as they have different NAVs. Additionally, the variable size could be improved by scholars. For this research, only the NAV was used to indicate the size of a fund. However, literature has signalled that the size of the industry and the size of the fund families can significantly influence the performance. It would be interesting to discover if these factors influence SRI mutual fund's performance. Extending the analysis to different periods would encourage further research to capture if different economic stages affect the results.

Similarly, this study should be extended to other countries, and the sample size could be increased. The results presented in this paper are only valid for a particular sample of funds for specific time periods. Lastly, future research should focus on the Shareholder engagement screen, including dialogue, shareholder resolutions, and proxy voting. The effort could contribute towards dividing these personal effects.

References

Alda, M. (2019). Corporate sustainability and institutional shareholders: The pressure of social responsible pension funds on environmental firm practices. *Business Strategy and the Environment*, 28(6), 1060-1071 https://doi-org.ezproxy2.utwente.nl/10.1002/bse.2301

Ambec, S., & Lanoie, P. (2008). Does it pay to be green? A systematic overview. *Academy of Management Perspectives*, 22(4), 45-62. DOI: 10.5465/AMP.2008.35590353

Auer, B., R. (2014). Do Socially Responsible Investment Policies Add or Destroy European Stock Portfolio Value? *Journal of Business Ethics*, 135, 381-397. https://doi-org.ezproxy2.utwente.nl/10.1007/s10551-014-2454-7

Barnett, M., L., Salomon, R., M. (2006). Beyond Dichotomy: The Curvilinear Relationship Between Social Responsibility and Financial Performance. *Strategic Management Journal*, 27(11), 1101-1122. https://www.jstor.org/stable/20142399

Bauer R, Otten R, Rad AT. 2006. Ethical investing in Australia: is there a financial penalty? *Pacific-Basin Finance Journal* 14, 33–48.

Baumol, W., Goldfeld, S., M., Gordon, L.A, Köhn, F-M. (1990). The economics of Mutual Fund Markets: Competition Versus Regulation. Kluwer Academic Publishers, Norwell, Massachusetts.

Brealey, R.A., Myers, R.C., Allen, F. (2009). Principles of Corporate Finance (13th ed.) New York: McGraw-Hill Education.

Buldyrev, S.V., Flori, A., Pammolli, F. (2021) Market instability and the size-variance relationship. Scientific Reports 11, 5737 https://doi.org/10.1038/s41598-021-84680-1

Camilleri, M.A. (2020), "The market for socially responsible investing: a review of the developments", *Social Responsibility Journal*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/SRJ-06-2019-0194

Capelle-Blancard, G., Monjon, S. (2017). The Performance of Socially Responsible Funds: Does the Screening Process Matter? *European Financial Management*, 20(3), 494-520 https://doi-org.ezproxy2.utwente.nl/10.1111/j.1468-036X.2012.00643.x

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

Chapple, L., Humphrey, J, E. (2013). Does Board Gender Diversity Have a Financial Impact? Evidence Using Stock Portfolio Performance. *Journal of Business Ethics*, 122, 709-723 https://doi-org.ezproxy2.utwente.nl/10.1007/s10551-013-1785-0

Cheng, B., Lounnou, L., Serafeim, G. (2013). Corporate social responsibility and access to finance. *Strategic management journal*, 35, 1-23.

Chiah, M., Chai, D., Zhong, A., Li, S. (2016). A Better Model? An Empirical Investigation of the Fama-French Five-factor Model in Australia. *International Review Of Finance*, 16 (4), 595-638. https://doi.org/10.1111/irfi.12099.

Cortez, M., C., Silva, F., Areal, N. (2009). The Performance of European Socially Responsible Funds. *Journal of Business Ethics*, 87(4), 573-588. DOI**:** 10.1007/s10551-008-9959-x

De Colle, S., York, J., G. (2009). Why Wine is not Glue? The Unresolved Problem of Negative Screening in Socially Responsible Investing. *Journal of Business Ethics*, 85, 83-95. DOI: 10.1007/s10551-008-9949-z

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

Droms, W., Walker, D. (1995). Mutual fund investment performance. *The Quarterly Review of Economics and Finance*, 36(3), 347-363.

Eccles, R., G., Ioannou, I., Serafeim, G. (2014). The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science*, 60(11), 2835-2857

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

El Ghoul, S., Karoui., A. (2017). Does corporate social responsibility affect mutual fund performance and flows? *Journal of Banking & Finance*, 77, 53-63. https://doi.org/10.1016/j.jbankfin.2016.10.009

Elton, E., Gruber MJ, Blake CR. 1996a. The persistence of risk-adjusted mutual fund performance. *Journal of Business,* 69(2), 133–157.

Erragragui, E., Lagoarde-Segot, T. (2016). Solving the SRI Puzzle? A note on the mainstreaming of ethical investment. *Finance Research Letters*, 18, 32-42. https://doi.org/10.1016/j.frl.2016.03.018

Eurosif (2018) SDGs for SRI investors. [Brochure] http://www.eurosif.org/wp-content/uploads/2018/01/Eurosif-SDGs-brochure.pdf

Fama, E. and French K. (1993). Common risk factors in the returns on stocks and bonds. *The journal of financial economics*, 33(1), 3-56.

Ferreira, M.A., Keswani, A., Miguel, A.F., Ramos, S.B. (2013). The Determinants of Mutual Fund Performance: A Cross-Country Study. *Review of Finance*, 17(2), 483-525 https://doi-org.ezproxy2.utwente.nl/10.1093/rof/rfs013

Global Sustainable Investment Alliance. (2018). Global Sustainable Investment Review. Retrieved from http://www.gsi-alliance.org/wp-content/uploads/2019/06/GSIR_Review2018F.pdf

Gompers, P., Ishii, J., Metrick, A. (2003). Corporate Governance and Equity Prices. *The Quarterly Journal of Economics*, 118(1), 107-156. https://doi.org/10.1162/00335530360535162

Haigh, M., Hazelton, J. (2004). Financial Markets: A Tool for Social Responsibility? *Journal of Business Ethics*, 52, 59-71.

Hair, J., F., Black, W., C., Babin B., J., Anderson, R., E. (2014). Multivariate Data Analysis (7th ed.) *Pearson New Internation Edition*.

Hayes, A. F., Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behaviour Research Methods*, 39(4), 709-722.

Hong, H., Kacperczyk, M., 2009. The price of sin: the effects of social norms on markets. *Journal of Financial Economics*, 93(1), 15-36 https://doi.org/10.1016/j.jfineco.2008.09.001 https://doi-org.ezproxy2.utwente.nl/10.1002/smj.2131

Humphrey, J. E., Lee, D. D. (2011). Australian Socially Responsible Funds: Performance, Risk and Screening Intensity. *Journal of Business Ethics*, 102(4), 519-535.

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

Johnsen, D, B. (2003). Socially Responsible Investing: A Critical Appraisal. *Journal of Business Ethics*, 43: 219-222

Jones, S., van der Laan, S., Frost, G., Loftus, J. (2007). The investment performance of Socially Responsible Investment Funds in Australia. *Journal of Business Ethics,* 80, 181-203. DOI 10.1007/s10551-007-9412-6

Jones, T., M., Harrison, J., S. (2018). How applying instrumental stakeholder theory can provide sustainable competitive advantage. *Academy of Management Review*, 43(3), 371-391. https://doi.org/10.5465/amr.2016.0111

Kacperczyk, M., Sialm, C., Zheng, L. (2004). On The Industry Concentration of Actively Managed Equity Mutual Funds. *National Bureau of Economic Research*. Working paper.

Kempf, A., Osthoff, P. (2007): The effect of socially responsible investing on portfolio performance, CFR Working Paper, No. 06-10, University of Cologne, Centre for Financial Research (CFR), Cologne

Khorana, A., Servaes, H., Tufano, P. (2004). Explaining the size of the mutual fund industry around the world. *Journal of Financial Economics*, 78, 145-185

Lee, D. D., Humphrey, J., Benson, K., L., Ahn, J., Y. (2010). Socially responsible investment fund performance: the impact of screening intensity. *Accounting & Finance*, 50(2), 351-370.

Lu, X., White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics*, 178 (1), 194-206. https://doi.org/10.1016/j.jeconom.2013.08.016
Margolis, J. D., Elfenbein, H. A., and Walsh, J. P. 2007. Does It Pay To Be Good? A Meta-Analysis and Redirection of Research on the Relationship between Corporate Social and Financial Performance, *Working paper*. Markets. *Journal of Financial Economics* 93, 5–36.
Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, 7, 77–91.
Martin, R., Ritchie, G. (2021, Oktober 7). *Can Your Pension Resits Climate Change U.K. Funds Tally Risk.* Bloomberg. https://www.bloomberg.com/news/articles/2021-10-07/can-your-pension-resist-climate-change-u-k-funds-tally-risks.
Pástor, L., Stambaugh, R., F., Taylor, L., A. (2015). Scale and skill in active management. *Journal of Financial Economics*. 116(1), 23-45
Pollet, J., M., Wilson, M. (2008). How Does Size Affect Mutual Fund Behavior? *Journal of Finance*, 63(6), 2941-2969
Pontiff, J. (1997). Excess Volatility and Closed-End Funds. *The American Economic Review*, 87(1), 155-169
Prather, L., Bertin, W. and Henker, T. (2004), “Mutual fund characteristics, managerial attributes, and fund performance”, *Review of Financial Economics*, 13(4), 305-326.
Renneboog, L., Ter Horst, J., Zhang, C. (2008a). Socially responsible investments: Institutional aspects, performance, and investor behaviour. *Journal of Banking & Finance*, 32(9), 1723-1742. https://doi.org/10.1016/j.jbankfin.2007.12.039
Renneboog, L., Ter Horst, J., Zhang, C. (2008b). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14(3), 302-322. https://doi.org/10.1016/j.jcorpfin.2008.03.009
Renneboog, L., Ter Horst, J., Zhang, C. (2011). Is ethical money financially smart? Non-financial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation*, 20(4), 562-588. https://doi.org/10.1016/j.jfi.2010.12.003
Revelli, C., Viviani, J-L., (2015). Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. Business Ethics: *A European Review*. 24(2) X
Scholtens, B. (2008). A note on the interaction between corporate social responsibility and financial performance. *Ecological Economics*, 68(1-2), 46-55. https://doi.org/10.1016/j.ecolecon.2008.01.024
Shakil, M., H. (2021). Environmental, social and governance performance and financial risk: Moderating role of ESG controversies and board gender diversity. *Resources policies* 72, 102144. https://doi.org/10.1016/j.resourpol.2021.102144
Singh, A.B. and Tandon, P. (2022), “Association between fund attributes and fund’s performance: a panel data approach”, *Benchmarking: An International Journal*, 29(1), 285-304. https://doi-org.ezproxy2.utwente.nl/10.1108/BIJ-10-2020-0545
Tirone, J. (2021, Oktober 7) *Blackrock Rainmaker Wants Wants Climate Cash for Emerging Markets*. Bloomberg. https://www.bloomberg.com/news/articles/2021-10-07/blackrock-s-hildebrand-says-emerging-markets-need-climate-boost.
Uysal, N., Yang, A., Taylor, M. (2017). Shareholder communication and issue salience: corporate responses to ‘social’ shareholder activism. *Journal of Applied Communication Research.* 46(2), 179-201. https://doi-org.ezproxy2.utwente.nl/10.1080/00909882.2018.1437643
Waddock, S., Smith, N. (2000). Corporate Responsibility Audits: Doing Well by Doing Good. *Sloan Management Review* 41(2), 75-83
Webster, D. (2002). Mutual Fund Performance and Fund Age. Available at SSRN: https://ssrn.com/abstract=1764543 or http://dx.doi.org/10.2139/ssrn.1764543

Appendices

Appendix I: Fees of mutual investment funds.

Table 1: Different fees mutual funds apply

The information on these types of fees are from the following website: https://www.investor.gov/introduction-investing/investing-basics/glossary/mutual-fund-fees-and-expenses

Name Fee	Describtion	subdivistion
Sales Loads	Compensation to the brokers. There are generally two types of sale loads: Front-end sales load that is payed when investors purchase the fund shares and back-end or deferred sales laod that is payed when investorers redeem their shares.	Shareholder Fees
Redemption Fee	This type of fee is also charged when a shareholder redeems their shares. The difference between a redemtion fee and a sales load is that the redemption fee is generally used to defray fund costs associated with a shareholder's redemption and is paid directly to the fund, not to a broker.	Shareholder Fees
Exchange Fee	Is a type of fee that some funds impose on shareholders if they transfer to another fund within the same fund group.	Shareholder Fees
Account Fee	Is a type of fee that some funds impose on shareholders for the maintenance of their account	Shareholder Fees
Purchase Fee	This type of fee is also charged when a shareholder purchages their shares. A redemption fee is paid to the fund and not to a broker and is generally imposed to defray some of the fund's costs associated with the purchase	Shareholder Fees
Management Fees	This fee is payed to the fund's investment advisor for managing the fund's investment portfolio and for administrative fees payable to the investment advisor.	Annual Fund Operating Expenses
Distribution (12b-1) Fees	These fees paid out of the fund assets to cover distribution expenses and sometimes shareholder service expenses	Annual Fund Operating Expenses
Total Annual Fund Operating Expenses	The total of a fund's annual fund operating expenes are expressed as a percentage of the fund's average net assets.	Annual Fund Operating Expenses
Other expenses.	This category include the expenses that are not included in the categories: management fee or distribution fees.	Annual Fund Operating Expenses

Appendix II: Table of screens used by Eurosif

Table 2: Screening strategies used by Eurosif

The information that is presented below can be found on de following website: http://www.eurosif.org/responsible-investment-strategies/

Screen category	Criteria
Best-in-class	"An approach where leading or best-performing investments within a universe, category, or class are selected or weighted based on ESG criteria."
Engagement & voting	"Engagement activities and active ownership through voting of shares and engagement with companies on ESG matters."
ESG integration	"The explicit inclusion by managers of ESG risks and opportunities into traditional financial analysis and investment decisions based on a systematic process."
Exclusions	"An approach that excluded specific investments or classes of investment."
Impact investing	"Impact Investments are investments made into companies, organisations and funds to generate social and environmental impact alongside a financial return."
Norms-based screening	"Screening of investments according to their compliance with international standards and norms."
Sustainability-themed	"Investment in themes or assets linked to the development of sustainability."

Appendix III: Comparable literature

Table 3: List of comparable literature

This table contains comparative literature used to determine the variables for this paper.

Researcher(s)	Titel research	Research methods	Dependent variable	Independent variables	control variables	reference
Barnett and Salomon (2006)	Beyond Dichotomy: The curvilinear relationship between social responsibility and financial performance	Ordinary least squares (OLS).	Risk-adjusted financial performance of SRI fund in month based on CAPM with Jensens alpha	Alcohol, Tobacco, Gabling, Defense/weapons, Animal Testing, Product/service quality, Environment, Human rights, Labour relations, Employment equality, Community investment, and Community relations	Fund age, total assets, global fund, percent stock, percent bonds, yearly dummy macro-economic factors.	Barnett, M., L., Salomon, R., M. (2006). Beyond Dichotomy: The Curvilinear Relationship Between Social Responsibility and Financial Performance. Strategic Management Journal, 27(11), 1101-1122.
Capelle-Blancard and Monjon (2012)	The Performance of Socially Responsible Funds: Does the Screening Process Matter?	Ordinary least squares (OLS).	Risk-adjusted financial performance over the whole period based on CAPM, Jensens alpha and 4-factors	Environment, Social, Governance	Fund age, Fund Size, Management Fees, Investment style (global, outside Europe, Europe), Bond, Balanced	Capelle-Blancard, G., Monjon, S. (2017). The Performance of Socially Responsible Funds: Does the Screening Process Matter? European Financial Management, Vol 20(3), 494-520
Lee, Humphrey, Ahn (2010)	Socially responsible investment fund performance: the impact of screening intensity	Ordinary least squares (OLS).	1) Jensens and Charhart abnormal return 2) 3 year annualised risk.	Alcohol, Tobacco, Gambling, Defense/weapons, Animal Testing, Product/service quality, Environment, Human rights, Labour relations, Employment equality, Community investment, and Community relations	Fund age, Fund Size, type (institutional or not), Proxy voting, allocation of assets, turnover.	Lee, D. D., Humphrey, J., Benson, K., L., Ahn, J., Y. (2010). Socially responsible investment fund performance: the impact of screening intensity. Accounting & Finance, Vol. 50, 2, pp, 351-370.
Renneboog, Ter Horst, Zhang (2008b)	The price of ethics and stakeholder governance: The performance of socially responsible mutual funds.	Ordinary least squares (OLS).	Risk-adjusted return in month based on CAPM, the Fama French Carhart (FFC) 4-factor model, and the expanded FFC model	SRI, Activism Policy, Community involvement, in-House SRI research, Islamic Fund, Ethical Screens, Social Screens Environment Screens.	Fund age, Fund size, Risk, Management Fees, Load Fees, Family size, Investing abroad, Country effects, Time effects	Renneboog, L., Ter Horst, J., Zhang, C. (2008b). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. Journal of Corporate Finance. 14(3), 302-322.

Appendix IV: Variable definitions

Table 4: Key variable definitions

The table below displays all the variables used for this paper.

Variable definitions

Name	Abbreviations	Measurement	References
Dependent variables			
Risk-Adjusted Return	RAP	CAPM, Fama French Three-factor, Carhart Four-factor, and Fama-French Five-factor	(Capelle-Blancard and Monjon, 2012; Lee et al., 2010; Renneboog et al., 2008b)
Independent variables			
Screening Category			
Environmental	D_Env	1 if fund i utilises Env criteria	(Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Social	D_Soc	1 if fund i utilises DSoc criteria	(Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Governance	D_Gov	1 if fund i utilises DGov criteria	(Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Product	D_Pro	1 if fund i utilises DPro criteria	(Barnett and Salomon, 2006; Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Other/Qualitative	D_Oth	1 if fund i utilises DOth criteria	(Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Shareholder Engagement	D_Sha	1 if fund i utilises Dsha criteria	Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Control variables			
Screening intensity, i	SI, i	Equals the number of screens used.	(Barnett and Salomon, 2006; Lee et al., 2010; Renneboog et al., 2008b)
	SI2	square of the number of criteria used	(Capelle-Blancard and Monjon, 2012)
Fund characteristic	FC, i, t-1		
Fund Size	Log_SIZE	The natural logarithm of fund assets under management in Dollar at month t-1	(Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Lee, Humphrey, Ahn, 2010; Renneboog et al., 2008b)
Fund Age	Log_AGE	The number of years since the fund's data of establishment	(Barnett and Salomon, 2006; Capelle-Blancard and Monjon, 2012; Humphrey and Lee, 2011; Lee et al., 2010; Renneboog et al., 2008b; Renneboog et al., 2011)
Global Fund	GLO	1 if the fund trades global, 0 otherwise	(Barnett and Salomon, 2006)
Expense Ratios	EXP	The lagged expense ratio	(Bauer et al., 2005; Pástor et al., 2015)

Appendix V: Sensitivity analysis

Table 5: Sensitivity analysis of screening criteria

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the OLS regression results on the relationship between the dependent variable RAP and the independent variable screening criteria. RAP, calculated with FF5, and the separate variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio (EXP) is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is invested internationally. All returns are monthly and in percentage.

Sensitivity analysis on screening criteria

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)	Model 7 (RAP)
(Constant)	0.014	0.012	0.011	0.012	0.015*	0.010	0.007
	(0.129)	(0.205)	(0.229)	(0.203)	(0.117)	(0.324)	(0.488)
D_Sha	-0.004						-0.027***
	(0.354)						(0.006)
D_Qua		0.002					0.006
		(0.655)					(0.174)
D_Pro			0.003				0.003
			(0.569)				(0.540)
D_Gov				0.000			-0.003
				(0.965)			(0.642)
D_Soc					-0.004		-0.004
					(0.219)		(0.251)
D_Env						0.003	0.033***
						(0.566)	(0.007)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002*
	(0.410)	(0.228)	(0.252)	(0.278)	(0.489)	(0.210)	(0.087)
EXP	0.003	0.004	0.001	0.003	0.003	0.003	0.001
	(0.521)	(0.455)	(0.800)	(0.489)	(0.545)	(0.486)	(0.914)
Log_AGE	-0.006*	-0.006*	-0.006*	-0.006*	-0.007**	-0.005*	-0.004
	(0.071)	(0.078)	(0.068)	(0.071)	(0.039)	(0.085)	(0.276)
GLO	0.003	0.003	0.004	0.003	0.003	0.003	0.003
	(0.305)	(0.298)	(0.247)	(0.331)	(0.305)	(0.346)	(0.369)
Adjusted R2	0.039	0.037	0.037	0.036	0.042	0.037	0.065
F- statistic (overall)	3.003**	2.862**	2.888**	2.820**	3.142***	2.890**	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 6: Step-by-step sensitivity analysis

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the OLS regression results on the relationship between the dependent variable RAP and the independent variable screening criteria. RAP, calculated with FF5, and the independent variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end on a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio (EXP) is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets are invested internationally. All returns are monthly and in percentage.

Sensitivity analysis on screening criteria

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)
(Constant)	0.014	0.014	0.013	0.013	0.016*	0.007
	(0.129)	(0.133)	(0.172)	(0.169)	(0.099)	(0.488)
D_Sha	-0.004	-0.005	-0.004	-0.006	-0.005	-0.027***
	(0.354)	(0.322)	(0.355)	(0.262)	(0.357)	(0.006)
D_Qua		0.002	0.002	0.002	0.005	0.006
		(0.568)	(0.504)	(0.612)	(0.191)	(0.174)
D_Pro			0.003	0.003	0.004	0.003
			(0.567)	(0.556)	(0.465)	(0.540)
D_Gov				0.003	0.005	-0.003
				(0.523)	(0.309)	(0.642)
D_Soc					-0.007*	-0.004
					(0.080)	(0.251)
D_Env						0.033***
						(0.007)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002*
	(0.410)	(0.330)	(0.296)	(0.277)	(0.281)	(0.087)
EXP	0.003	0.003	0.002	0.001	0.001	0.001
	(0.521)	(0.476)	(0.779)	(0.808)	(0.860)	(0.914)
Log_AGE	-0.006*	-0.005*	-0.005*	-0.006*	-0.007**	-0.004
	(0.071)	(0.081)	(0.081)	(0.077)	(0.029)	(0.276)
GLO	0.003	0.004	0.005	0.004	0.004	0.003
	(0.305)	(0.285)	(0.226)	(0.310)	(0.283)	(0.369)
Adjusted R2	0.039	0.037	0.034	0.031	0.040	0.065
F- statistic (overall)	3.003**	2.549**	2.226**	1.994**	2.132**	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 7: Sensitivity analysis grouped screening criteria

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the OLS regression results on the relationship between the dependent variable RAP and the independent variable screening criteria. RAP, calculated with FF5, and the separate variable screening category. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is measured in years from inception. The expense ratio (EXP) is the total of annual costs associated with investing in the fund expressed as a percentage of the investment. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is invested internationally. All returns are monthly and in percentage.

Sensitivity analysis on screening criteria

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)
(Constant)	0.011	0.016	0.011
	(0.264)	(0.107)	(0.261)
D_Sha		-0.006	
		(0.293)	
D_Qua	0.006		0.002
	(0.142)		(0.575)
D_Pro	0.005	0.003	0.004
	(0.401)	(0.630)	(0.504)
D_Gov		0.006	0.000
		(0.262)	-0.984
D_Soc	-0.007*	-0.004	
	(0.053)	(0.204)	
D_Env	0.005		
	(321)		
Log_SIZE	-0.002	-0.001	-0.001
	(0.172)	(0.478)	(0.211)
EXP	0.001	0.001	0.001
	(0.864)	(0.857)	(0.799)
Log_AGE	-0.007**	-0.007**	-0.006*
	(0.037)	(0.034)	(0.079)
GLO	0.005	0.003	0.005
	(0.184)	(0.407)	(0.224)
Adjusted R2	0.043	0.037	0.030
F- statistic (overall)	2.374**	2.176**	2.096**

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Appendix VI: The Robust regression

Table 8: Robust regression on the impact of screening intensity on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Parameter Estimates with Robust Standard Errors

	Model 1 (RAP)	Model 2 (RAP)
Intercept	0.013	-0.049
	(0.298)	(0.412)
SI	-9E-05	0.009
	(0.907)	(0.299)
SI2		0.000
		(0.287)
Log_SIZE	-0.001	-0.001
	(0.454)	(0.487)
EXP	0.003	0.004
	(0.453)	(0.401)
Log_AGE	-0.006*	-0.006*
	(0.098)	(0.093)
GLO	0.003	0.003
	(0.314)	(0.395)
Adjusted R2	0.036	0.037
F-statistic (overall)	2.823**	2.550**

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 9: Robust regression on the impact of screening criteria on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)	Model 7 (RAP)
(Constant)	0.010	0.015	0.012	0.011	0.012	0.014	0.007
	(0.263)	(0.180)	(0.140)	(0.264)	(0.168)	(0.102)	(0.473)
D_Env	0.003						0.033**
	(0.605)						(0.013)
D_Soc		-0.004					-0.004
		(0.225)					(0.259)
D_Gov			0.000				-0.003
			(0.964)				(0.487)
D_Pro				0.003			0.003
				(0.509)			(0.483)
D_Qua					0.002		0.006
					(0.673)		(0.211)
D_Sha						-0.004	-0.027**
						(0.443)	(0.023)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002
	(0.309)	(0.573)	(0.379)	(0.439)	(0.330)	(0.508)	(0.177)
EXP	0.003	0.003	0.003	0.001	0.004	0.003	0.001
	(0.453)	(0.518)	(0.455)	(0.858)	(0.425)	(0.486)	(0.909)
Log_AGE	-0.005	-0.007*	-0.006*	-0.007*	-0.006	-0.006*	-0.004
	(0.111)	(0.059)	(0.100)	(0.080)	(0.109)	(0.099)	(0.322)
GLO	0.003	0.003	0.003	0.004	0.003	0.003	0.003
	(0.346)	(0.309)	(0.347)	(0.342)	(0.292)	(0.306)	(0.422)
Adjusted R2	0.037	0.037	0.036	0.037	0.037	0.039	0.065
F- statistic (overall)	2.890**	2.890**	2.820**	2.888**	2.862**	3.003**	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 10: Robust regression on the step-by-step inclusion of screening criteria.

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)	Model 4 (RAP)	Model 5 (RAP)	Model 6 (RAP)
(Constant)	0.010	0.012	0.012	0.011	0.011	0.007
	(0.263)	(0.180)	(0.209)	(0.264)	(0.263)	(0.473)
D_Env	0.003	0.005	0.005	0.005	0.006	0.033**
	(0.605)	(0.420)	(0.422)	(0.418)	(0.344)	(0.013)
D_Soc		-0.004	-0.004	-0.004	-0.007*	-0.004
		(0.159)	(0.200)	(0.203)	(0.077)	(0.259)
D_Gov			0.000	0.000	0.000	-0.003
			(0.983)	(0.937)	(0.919)	(0.487)
D_Pro				0.003	0.005	0.003
				(0.509)	(0.336)	(0.483)
D_Qua					0.006	0.006
					(0.172)	(0.211)
D_Sha						-0.027**
						(0.023)
Log_SIZE	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002
	(0.309)	(0.460)	(0.460)	(0.439)	(0.280)	(0.177)
EXP	0.003	0.003	0.003	0.001	0.001	0.001
	(0.453)	(0.525)	(0.526)	(0.858)	(0.853)	(0.909)
Log_AGE	-0.005	-0.006*	-0.006*	-0.007*	-0.007*	-0.004
	(0.111)	(0.066)	(0.082)	(0.080)	(0.078)	(0.322)
GLO	0.003	0.003	0.003	0.004	0.005	0.003
	(0.346)	(0.359)	(0.384)	(0.342)	(0.242)	(0.422)
Adjusted R2	0.037	0.041	0.037	0.034	0.039	0.065
F- statistic (overall)	2.890**	2.746**	2.344**	2.084**	2.102**	2.709***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 11: Robust regression for the impact of different groups of screening criteria

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF5, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total of annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage

Variables	Model 1 (RAP)	Model 2 (RAP)	Model 3 (RAP)
(Constant)	0.011	0.016*	0.011
	(0.231)	(0.093)	(0.205)
D_Env	0.005		
	(0.376)		
D_Soc	-0.007*	-0.004	
	(0.059)	(0.205)	
D_Gov		0.006	0.000
		(0.212)	(0.985)
D_Pro	0.005	0.003	0.004
	(0.332)	(0.570)	(0.426)
D_Qua	0.006		0.002
	(0.171)		(0.598)
D_Sha		-0.006	
		(0.313)	
Log_SIZE	-0.002	-0.001	-0.001
	(0.280)	(0.569)	(0.316)
EXP	0.001	0.001	0.001
	(0.855)	(0.854)	(0.781)
Log_AGE	-0.007*	-0.007*	-0.006
	(0.059)	(0.057)	(0.113)
GLO	0.005	0.003	0.005
	(0.234)	(0.450)	(0.269)
Adjusted R2	0.043	0.037	0.030
F- statistic (overall)	2.374**	2.176**	2.096**

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Appendix VII: Robustness check using the CAPM model

Table 12: OLS regression on the impact of screening intensity on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CAPM, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

	Model 1 (CAPM)	Model 2 (CAPM)
Constant	-0.059	-0.103
	(0.203)	(0.610)
SI	0.000	0.006
	(0.851)	(0.835)
SI2		0.000
		(0.822)
Log_SIZE	0.000	0.000
	(0.925)	(0.108)
EXP	-0.010	-0.010
	(0.542)	(0.555)
Log_AGE	-0.001	-0.002
	(0.891)	(0.887)
GLO	0.008	0.008
	(0.455)	(0.483)
Adjusted R2	-0.018	-0.022
F-statistic (overall)	0.143	0.127

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table: 13: OLS regression on the impact of screening category on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CAPM, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (CAPM)	Model 2 (CAPM)	Model 3 (CAPM)	Model 4 (CAPM)	Model 5 (CAPM)	Model 6 (CAPM)	Model 7 (CAPM)
(Constant)	-0.066**	-0.064**	-0.064**	-0.070**	-0.064**	-0.060*	-0.078**
	(0.048)	(0.045)	(0.040)	(0.026)	(0.038)	(0.063)	(0.029)
D_Env	0.001						0.048
	(0.954)						(0.256)
D_Soc		-0.002					0.004
		(0.831)					(0.757)
D_Gov			-0.002				-0.007
			(0.857)				(0.742)
D_Pro				0.017			0.014
				(0.357)			(0.461)
D_Qua					-0.005		-0.004
					(0.639)		(0.779)
D_Sha						-0.008	-0.021
						(0.587)	(0.223)
Log_SIZE	0.00	0.000	0.000	0.000	0.001	0.001	0.000
	(0.991)	(0.940)	(0.950)	(0.975)	(0.842)	(0.873)	(0.902)
EXP	-0.010	-0.010	-0.010	-0.020	-0.011	-0.011	-0.021
	(0.537)	(0.528)	(0.535)	(0.303)	(0.497)	(0.518)	(0.293)
Log_AGE	-0.001	-0.002	-0.001	-0.001	-0.002	-0.001	0.003
	(0.922)	(0.875)	(0.918)	(0.879)	(0.882)	(0.917)	(0.774)
GLO	0.008	0.008	0.009	0.014	0.008	0.009	0.011
	(0.458)	(0.451)	(0.439)	(0.273)	(0.477)	(0.446)	(0.409)
Adjusted R2	-0.018	-0.018	-0.018	-0.014	-0.017	-0.017	-0.029
F- statistic (overall)	0.14	0.145	0.142	0.307	0.180	0.195	0.314

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 14: **OLS regression for the impact of different groups of screening criteria**

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CAPM, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (CAPM)	Model 2 (CAPM)	Model 3 (CAPM)
(Constant)	-0.071**	-0.065*	-0.069**
	(0.040)	(0.054)	(0.032)
D_Env	0.003		
	(0.857)		
D_Soc	0.000	-0.001	
	(0.977)	(0.925)	
D_Gov		0.005	0.000
		(0.799)	(0.922)
D_Pro	0.016	0.016	0.015
	(0.395)	(0.386)	(0.402)
D_Qua	-0.003		-0.003
	(0.810)		(0.799)
D_Sha		-0.009	
		(0.634)	
Log_SIZE	0.000	0.000	0.000
	(0.973)	(0.946)	(0.930)
EXP	-0.020	-0.020	-0.020
	(0.302)	(0.305)	(0.306)
Log_AGE	-0.001	-0.002	-0.002
	(0.893)	(0.879)	(0.877)
GLO	0.013	0.013	0.013
	(0.311)	(0.324)	(0.314)
Adjusted R2	-0.025	-0.026	-0.023
F- statistic (overall)	0.267	0.221	0.229

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Appendix VIII: Robustness check using Fama-French three-factor

Table 15: OLS regression on the impact of screening intensity on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF3, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (FF3)	Model 2 (FF3)
Constant	-0.054	-0.129
	(0.244)	(0.537)
SI	-0.001	0.010
	(0.618)	(0.748)
SI2		0.000
		(0.714)
Log_SIZE	0.002	0.002
	(0.675)	(0.660)
EXP	-0.005	-0.005
	(0.762)	(0.784)
Log_AGE	-0.003	-0.003
	(0.789)	(0.782)
GLO	0.009	0.008
	(0.416)	(0.455)
Adjusted R2	-0.017	-0.020
F-statistic (overall)	0.197	0.186

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 16: OLS regression on the impact of screening category on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF3, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (FF3)	Model 2 (FF3)	Model 3 (FF3)	Model 4 (FF3)	Model 5 (FF3)	Model 6 (FF3)	Model 7 (FF3)
(Constant)	-0.07**	-0.067**	-0.070**	-0.077**	-0.071**	-0.062*	-0.081**
	(0.044)	(0.043)	(0.031)	(0.017)	(0.027)	(0.062)	(0.027)
D_Env	-0.003						0.059
	(0.850)						(0.180)
D_Soc		-0.007					0.00
		(0.511)					(0.998)
D_Gov			-0.005				-0.006
			(0.700)				(0.776)
D_Pro				0.019			0.015
				(0.303)			(0.425)
D_Qua					-0.007		-0.003
					(0.536)		(0.848)
D_Sha						-0.016	-0.056
						(0.327)	(0.113)
Log_SIZE	0.001	0.002	0.001	-0.001	0.002	0.002	0.001
	(0.788)	(0.684)	(0.750)	(0.866)	(0.654)	(0.632)	(0.893)
EXP	-0.005	-0.006	-0.006	-0.017	-0.007	-0.006	-0.018
	(0.745)	(0.713)	(0.742)	(0.397)	(0.684)	(0.707)	(0.375)
Log_AGE	-0.002	-0.004	-0.002	-0.002	-0.003	-0.002	0.002
	(0.835)	(0.727)	(0.858)	(0.831)	(0.809)	(0.855)	(0.851)
GLO	0.010	0.004	0.011	0.016	0.009	0.010	0.012
	(0.399)	(0.403)	(0.371)	(0.224)	(0.438)	(0.397)	(0.377)
Adjusted R2	-0.018	-0.016	-0.017	-0.013	-0.016	-0.014	-0.022
F- statistic (overall)	0.14	0.234	0.177	0.361	0.224	0.340	0.467

* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Table 17: OLS regression for the impact of different groups of screening category

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with FF3, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (FF3)	Model 2 (FF3)	Model 3 (FF3)
(Constant)	-0.073**	-0.065*	-0.075**
	(0.042)	(0.060)	(0.022)
D_Env	0.001		
	(0.965)		
D_Soc	-0.005	-0.005	
	(0.673)	(0.647)	
D_Gov		0.008	-0.002
		(0.665)	(0.876)
D_Pro	0.018	0.017	0.017
	(0.350)	(0.358)	(0.402)
D_Qua	-0.002		-0.003
	(0.900)		(0.714)
D_Sha		-0.017	
		(0.405)	
Log_SIZE	0.001	0.002	0.001
	(0.739)	(0.668)	(0.726)
EXP	-0.017	-0.017	-0.017
	(0.395)	(0.393)	(0.405)
Log_AGE	-0.004	-0.004	-0.003
	(0.733)	(0.738)	(0.810)
GLO	0.015	0.014	0.015
	(0.252)	(0.298)	(0.258)
Adjusted R2	-0.025	-0.022	-0.021
F- statistic (overall)	0.267	0.353	0.284

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Appendix IX: Robust regression using CH4

Table 18: OLS regression for the impact of screening intensity RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CAPM, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. Screening intensity is measured as the absolute number of screens used by each fund. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (CH4)	Model 2 (CH4)
Constant	-0.021	-0.075
	(0.653)	(0.715)
SI	-0.001	0.007
	(0.724)	(0.812)
SI2		0.000
		(0.787)
Log_SIZE	0.003	0.003
	(0.438)	(0.430)
EXP	0.039**	0.039**
	(0.020)	(0.020)
Log_AGE	-0.036***	-0.036***
	(0.001)	(0.001)
GLO	-0.007	-0.008
	(0.529)	(0.508)
Adjusted R2	0.059	-0.020
F-statistic (overall)	4.064***	3.386***

* Statistical significance at the 10% level
** Statistical significance at the 5% level
*** Statistical significance at the 1% level

Table 19: OLS regression on the impact of screening criteria on RAP

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CH4, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (CH4)	Model 2 (CH4)	Model 3 (CH4)	Model 4 (CH4)	Model 5 (CH4)	Model 6 (CH4)	Model 7 (CH4)
(Constant)	-0.032	-0.028	-0.035	-0.029	-0.034	-0.031	-0.017
	(0.348)	(0.381)	(0.268)	(0.360)	(0.276)	(0.350)	(0.639)
D_Env	-0.002						-0.008
	(0.892)						(0.854)
D_Soc		-0.007					-0.014
		(0.519)					(0.313)
D_Gov			0.005				0.015
			(0.748)				(0.475)
D_Pro				-0.018			-0.016
				(0.329)			(0.394)
D_Qua					0.004		0.008
					(0.727)		(0.578)
D_Sha						-0.005	-0.005
						(0.773)	(0.878)
Log_SIZE	0.003	0.003	0.002	0.003	0.002	0.003	0.003
	(0.488)	(0.392)	(0.571)	(0.462)	(0.624)	(0.475)	(0.500)
EXP	0.038**	0.038**	0.039**	0.049**	0.039**	0.038**	0.048**
	(0.021)	(0.024)	(0.021)	(0.014)	(0.020)	(0.022)	(0.017)
Log_AGE	-0.036***	-0.037***	-0.036***	-0.035***	-0.035***	-0.036***	-0.039***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
GLO	-0.007	-0.007	-0.008	-0.013	-0.007	-0.007	-0.014
	(0.547)	(0.540)	(0.499)	(0.314)	(0.556)	(0.539)	(0.309)
Adjusted R2	0.058	0.060	0.059	0.062	0.059	0.059	0.048
F- statistic (overall)	4.042***	4.128***	4.060***	4.245***	4.064***	4.056***	2.241**

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level

Table 20: OLS regression for the impact of different groups of screening criteria

The dataset ranges from January 1st, 2015, until December 31st, 2020 and is balanced. The SRI sample consists of 41 funds. This table presents the robust regression results on the relationship between the dependent variable RAP, calculated with CH4, and the independent variable screening intensity. RAP defines as the average monthly return, measured as the percentage change in a fund's market value from the beginning to the end of a 12-month estimate, adjusted by specific risk factors. The screening categories are assigned a dummy of 1 if the fund screens for all screening criteria in the particular category. Fund age is in years from inception. The expense ratio (EXP) is the total annual costs associated with investing in the fund expressed as a percentage of the investment and lagged. Log size is the natural logarithm of the total assets under management. The global dummy (GLO) signals if part of the fund's assets is international. Screening intensity squared is added in model two to check for a non-linear relationship. All returns are monthly and in percentage.

Variables	Model 1 (CH4)	Model 2 (CH4)	Model 3 (CH4)
(Constant)	-0.021	-0.020	-0.030
	(0.544)	(0.564)	(0.349)
D_Env	-0.001		
	(0.959)		
D_Soc	-0.011	-0.009	
	(0.363)	(0.408)	
D_Gov		0.014	0.002
		(0.443)	(0.871)
D_Pro	-0.017	-0.018	-0.017
	(0.377)	(0.320)	(0.356)
D_Qua	0.009		0.001
	(0.530)		(0.913)
D_Sha		-0.012	
		(0.555)	
Log_SIZE	0.003	0.003	0.002
	(0.505)	(0.387)	(0.576)
EXP	0.049**	0.048**	0.049**
	(0.015)	(0.017)	(0.015)
Log_AGE	-0.038***	-0.038***	-0.035***
	(0.001)	(0.001)	(0.001)
GLO	-0.012	-0.016	-0.013
	(0.377)	(0.241)	(0.330)
Adjusted R2	0.054	0.055	0.054
F- statistic (overall)	2.745***	2.777***	3.015***

* Statistical significance at the 10% level

** Statistical significance at the 5% level

*** Statistical significance at the 1% level