The Performance of Socially Responsible Investments: A Meta-Analysis

Lars Hornuf
University of Bremen
Chair of Business Administration,
Financial Services, and Financial Technology
Max-von-Laue-Straße 1, 28359 Bremen, Germany

phone: +49 (0)421 218 - 66 820
email: hornuf@uni-bremen.de
website: www.hornuf.com

Gül Yüksel
University of Bremen
Max-von-Laue-Straße 1, 28359 Bremen, Germany

phone: +49 (0)421 218 - 66 826
email: guel.yueksel@uni-bremen.de

Acknowledgments

This article evolved as part of the research project "Transformation-oriented innovation policy in the bio-economy - status, potentials and strategies," which has been supported by the Federal Ministry of Education and Research under grant number 031B0781A. The authors thank Gregor Dorfleitner, Christian Fieberg, Sebastian Gehricke, Fabian Hollstein, Bing Liang, Luc Renneboog, Simon Rottke, Eliza Stenzhorn, Armin Varmaz, Karl Weinmayer and the participants of the 8th Paris Financial Management Conference (IPAG Business School), SFiC Annual Conference Sustainability, Climate Change and Financial Innovation (University of Birmingham Dubai), the 3rd CEFGroup Climate Finance Symposium (University of Otago), the TILEC Seminar (Tilburg University), IFABS 2022 Naples Conference (University of Naples "Federico II"), the 15th Edition of the Annual Meeting of The Risk, Banking and Finance Society (LUM Giuseppe Degennaro—Free Mediterranean University), the HVB Doctoral seminar (University of Bremen), and the Conference on Sustainable Finance and Digitalization (Bremen University of Applied Sciences and DIW Berlin) for their helpful comments and suggestions.

The Performance of Socially Responsible Investments: A Meta-Analysis

Abstract

In this article, we use a meta-analysis to examine the performance of socially responsible investing (SRI). After a thorough literature search, we review 153 empirical studies containing 1,047 observations of SRI performance. We find that, on average, SRI neither outperforms nor underperforms the market portfolio. However, in line with modern portfolio theory, we find that global SRI portfolios outperform regional sub-portfolios. Moreover, high-quality publications, publications in finance journals, and authors who publish more frequently on SRI are all less likely to report SRI outperformance. In particular, we find that including more factors in a capital market model reduces the likelihood that a study will find SRI outperformance. These findings have important implications for the policy evaluation of environmental, social, and governance goals in general, the asset management literature in particular, and the perspective of different scientific disciplines.

This version: January 5, 2023

JEL Classification: G11, G12, M14

Keywords: environmental social governance; ESG; socially responsible investment; SRI; meta-analysis

1 Introduction

Environmental, social, and governance (ESG) criteria have gained increasing importance in any field of the economy during the last decade. However, the extent to which investors benefit or lose by investing in ESG compliant assets is still disputed. Some scholars argue that socially responsible investing (SRI) that incorporates ESG criteria not only has positive environmental impacts but also can generate excessive financial returns (Derwall et al., 2011). However, this claim is at odds with modern portfolio theory, which states that investors who wish to maximize their risk-adjusted returns should select assets from the greatest possible universe and thus should not exclude assets scoring low on ESG criteria. The empirical evidence on that question is still mixed. For example, early on Renneboog et al. (2008) investigated funds across the world and found that SRI funds in the US, the UK, and in many continental European and Asia-Pacific countries underperform their domestic benchmarks. By contrast, risk-adjusted returns of SRI funds were not statistically different from the performance of conventional funds for most countries, with the exception of France, Ireland, Japan and Sweden, where SRI funds underperformed. In a similar vein, a meta-analysis by Revelli and Viviani (2015, p. 158) summarizes that SRI is "neither a weakness nor a strength compared with conventional investments."

In this article, we methodically examine the results of different types of performance assessments that have been undertaken in the literature. We conduct a systematic literature review and identify 5,845 studies related to ESG criteria and SRI. Overall, 153 of these studies have conducted a primary empirical study, the results of which we consider in a meta-analysis. In particular, we are interested in how primary study characteristics such as the quality of the journal or the methods applied affect the SRI performance found by the authors. While we are not the first to conduct a meta-analysis on SRI performance (see, e.g., Friede et al., 2015; Horváthová, 2010; Rathner, 2013), the tendency to invest responsibly has gained momentum especially in the last five years, not least because the EU adopted a taxonomy for sustainable activities, which entered into force on July 12, 2020. Between 2016 and 2020, global ESG integration had already increased by 143%, and total assets under ESG management today account for USD 25.2 trillion (Global Sustainable Investment Alliance, 2021). This shift toward ESG assets under management may have affected not only their performance as such but also the findings of empirical research on SRI performance evaluations. Systematically evaluating the current empirical evidence is therefore essential.

We contribute to the literature in at least three ways. First, we conduct what we believe is the most sophisticated meta-analysis of SRI performance, including the most recent publications. This analysis, taking into account the latest available literature, also enables us to take into account the current increase in demand for ESG assets. Second, our analysis is also broader, in that we include studies that use a wide variety of methods, datasets, and regions from various sub-fields of the business, economics, entrepreneurship, and finance literature. The dataset of SRI studies we have created enables us also to consider the journal type and quality and the complexity of the capital market models used when analyzing SRI performance. Third, we indirectly contribute to the recent literature on the newness and replicability of results (Serra-

Garcia & Gneezy, 2021), by showing that some authors and sub-fields in the literature are more consistent in their theoretical predictions and empirical results regarding SRI performance, while other sub-fields come to more extraordinary empirical results; SRI not only performs as well as the market portfolio but outperforms it.

In line with previous meta-studies (Revelli & Viviani, 2015; Wallis & Klein, 2015), we find that, on average, SRI neither outperforms nor underperforms the market portfolio. Nevertheless, we find that global SRI portfolios outperform regional sub-portfolios, which is in line with the postulates of modern portfolio theory. Furthermore, high-quality publications, publications in finance journals, and authors who publish more frequently on SRI are all less likely to report SRI outperformance. In particular, we find that including more factors in a capital market model reduces the likelihood that SRI outperformance is found in a study. These findings have important implications not only for the policy evaluation of ESG goals in general but also for the asset management literature in particular.

The remainder of the paper proceeds as follows: in Section 2, we describe the relevant theory and testable hypotheses. Section 3 presents the data and method. Section 4 outlines the empirical results and presents robustness checks. Section 5 concludes.

2 Hypotheses

Modern portfolio theory predicts that investors benefit from diversification, because diversification increases the risk-adjusted returns if assets are not perfectly correlated (Markowitz, 1952). Fund managers who wish to maximize their risk-adjusted returns should therefore choose assets from the greatest possible universe. Because SRI excludes certain assets that do not meet the strict sustainability criteria, socially responsible fund managers are faced with a smaller universe of assets that can be included in the respective portfolio. Consequently, fund managers focusing on SRI might not build efficient mean-variance portfolios and, in line with modern portfolio theory, should expect risk-adjusted returns that are inferior to those of conventional fund managers if the available universe of assets becomes too small to construct portfolios at the efficiency frontier.

While modern portfolio theory posits how investors should behave, the question is how investors actually behave. Bauer et al. (2021) show that some pension fund members indeed vote for a more sustainable investment policy out of social preferences rather than higher expected returns. Furthermore, Renneboog et al. (2011) point out that ethical investors are less concerned about negative returns than conventional investors because ethical investments flows are less sensitive to past negative returns than conventional investments flows. Ethical investors are interested in non-financial characteristics of mutual funds that are linked to investors' ethical or social preferences. In addition to building suboptimal portfolios, the screening for socially responsible assets also increases search and monitoring costs, which further reduces the investment performance after transaction costs (Bauer et al., 2006). These arguments represent the so-called *underperformance hypothesis* of the SRI literature, according to which investors, in return for supporting companies that attach great importance to ESG criteria, need to expect and therefore accept portfolios with lower returns than the market portfolio (Badía et al., 2020; Chan & Walter, 2014; Nainggolan et al., 2016; Pastor et al., 2021b; Renneboog et al., 2008).

Finance scholars focusing on portfolio management have initially examined the underperformance hypothesis using more refined theoretical models. Heinkel et al. (2001) were among the first to develop an equilibrium model showing that non-green firms confront higher costs of capital, because green investors avoid these firms and overall fewer investors demand non-green assets. Consequently, investors who do not have ESG preferences earn a premium because others avoid non-green firms. More recently, Luo and Balvers (2017) and Zerbib (2020) confirmed that, in equilibrium, segmented markets lead to higher expected returns for investors who include non-green firms in their portfolios. Thus, in theory, investors who have only risk-and-return motives should also hold non-ESG assets and consider so-called sin stocks, because doing so increases the diversification potential of their portfolio (Renneboog et al., 2008).

The results of empirical research that analyzes whether ESG portfolios underperform the market are mixed. In support of the underperformance hypothesis, Dorfleitner and Grebler (2022) find that firms with high corporate social performance have a lower systematic risk, but also a lower stock market performance, because they participate less in the positive overall market development in the long term. Hudson (2005) and Pasewark and Riley (2010) point out that investors with ESG preferences cannot realize them as ESG funds come with higher costs, such as the additional expenses for fund management and monitoring the ESG ranking of companies, which investors are not willing to bear. The authors stress that socially responsible investors must accept lower returns to follow their principles. Pedersen et al. (2021) recently provided nuanced empirical evidence that ESG preferences are costly if investors incorporate such criteria into their portfolios. When the authors measure ESG accounting decisions against provisions in the financial statement, the maximum Sharpe ratio is reached for a comparatively high ESG level. If the ESG level continues to rise, the Sharpe ratio falls only slightly and ESG preferences can be realized at low cost. However, when imposing more realistic constraints on ESG portfolios, a steeper reduction in the ESG Sharpe ratio results. Finally, when removing assets with extremely low ESG ratings, green investors who maximize their Sharpe ratio might choose portfolios with even lower ESG portfolio scores than unconstrained investors, who can also invest in these low-ESG score assets. This is because unconstrained investors can short low ESG score assets to hedge risks, which constrained investors cannot. Relatedly, Barber et al. (2021) find that investors in impact funds realize 2.5 to 3.7 percentage points lower internal rates of return than investors in traditional venture capital funds. Larcker and Watts (2020) investigate whether individuals who support environmentally friendly projects are willing to give up higher returns. They observe that the pricing differential between a sample of green bonds and a matched sample of conventional bonds is very small, indicating the absence of a green pricing premium.

The contrasting hypothesis to the underperformance hypothesis is the *outperformance hypothesis*. This hypothesis, postulated primarily by corporate governance and corporate finance scholars, suggests that investors focusing on ESG criteria should expect excess returns, because firms scoring high on ESG criteria are better equipped to deal with environmental and social crises (Badía et al., 2020; Chan & Walter, 2014; Nainggolan et al., 2016; Renneboog et al., 2008). Intensive managerial screening of ESG criteria can reduce the costs that a company may incur as a result of environmental and social crises. If financial markets underestimate the cost of an environmental and social crisis, a company with high ESG scores may later outperform its benchmark (Edmans, 2011). Put differently, strong social and environmental performance is a sign of top management quality that translates into positive financial performance (Renneboog et al., 2011). This is in line with findings by Cornell (2021), indicating that companies scoring high on ESG criteria can reduce their cost of capital. Furthermore, Pastor et al. (2021a) show that high realized returns on green assets also result from news media reporting information about environmental concerns.

More generally, increased demand by investors with ESG preferences can—at least in the short run—drive SRI outperformance. Because of increasing awareness for environmental topics and preferences shifting toward SRI in recent years, assets scoring high on ESG criteria have shown positive abnormal returns. While the annual performance of the MSCI World SRI Index[1] rose by 12.12 percentage points between 2016 and 2020, the MSCI World Index[2] gained only 8.35 percentage points during this period. Especially in Europe, assets managed according to SRI strategies have increased strongly, with a corresponding positive impact on stock prices (Mollet et al., 2013). According to a 2021 report by the European Fund and Asset Management Association, almost EUR 11 trillion in assets are managed in Europe by means of an ESG investment approach. The annual performance of the MSCI Europe SRI Index in 2021 was 28.20%, which corresponds to an outperformance of the conventional MSCI Europe Index by 2.35 percentage points[3].

Theoretical research does not predict that stocks scoring low on ESG criteria should outperform the market, even though market portfolios also including non-ESG stocks should outperform pure ESG market portfolios. Hong and Kacperczyk (2009) find that over the period 1962–2006, sin stocks trading on the NYSE, Amex, and Nasdaq realized positive abnormal returns, largely because investors viewed these stocks as riskier. Bolton and Kacperczyk (2021) evidence that institutional investors screen assets for their emission intensity and that firms with higher total carbon dioxide emissions have earned higher returns. By contrast, Gompers et al. (2003) and Edmans (2011) find evidence that higher scores on certain ESG criteria, such as worker satisfaction, correlate with higher firm profits. The products of companies with high ESG scores allow higher profit margins, because the demand is less price elastic (Albuquerque et al., 2019). Servaes and Tamayo (2013) show that corporate social responsibility (CSR) activities of firms with high public and customer awareness have therefore increased in firm value while CSR concerns can cause greater harm to these firms. Furthermore, an early meta-analysis

---

[1] MSCI World SRI Index 2022. Retrieved February 11, 2022, from https://www.msci.com/documents/10199/641712d5-6435-4b2d-9abb-84a53f6c00e4.
[2] MSCI World Index 2022. Retrieved February 11, 2022, from https://www.msci.com/documents/10199/178e6643-6ae6-47b9-82be-e1fc565ededb.
[3] MSCI Europe SRI Index 2022. Retrieved February 11, 2022, from https://www.msci.com/documents/10199/19bb57cc-7077-4ab1-a4de-b3e51afa4be6.

investigating more than 2000 observations on ESG performance concludes that the majority of studies find a positive ESG impact on corporate financial performance, which is robust over time. The outperformance of ESG assets is found primarily in articles examining North America, emerging markets, and non-equity assets (Friede et al., 2015).

In summary, at least in the short run, additional demand for SRI, shifting environmental preferences, and news about environmental disasters can result in comparatively high expected returns for green assets. However, in the long run, the expected returns of assets scoring high on ESG criteria should—in line with modern portfolio theory—be lower than those of the market portfolio, because diversification increases risk-adjusted returns and ESG criteria might make it impossible to beat a market benchmark. Our meta-analysis intends to systematically analyze the long-term returns of assets scoring high on ESG criteria. We therefore follow the underperformance hypothesis and hypothesize:

H1: Socially responsible investments underperform the market portfolio.

Investors can reduce firm-specific risk not only through the diversification of assets but also through the diversification of assets from different regions. In line with modern portfolio theory, investors benefit from international investments (Grauer & Hakansson, 1987; Grubel, 1968; Lessard, 1976; Markowitz, 1952; Solnik, 1974) and, in general, suffer from a local bias (Coval & Moskowitz, 1999; French & Poterba, 1991; Ivković & Weisbenner, 2005). For example, Eun et al. (1991) compare US investments with international investments and show that the latter outperforms the former. The Sharpe ratio of the internationally diversified portfolio is higher by 0.051 than that of the S&P 500 Index. Treynor and Jensen's measures of the internationally diversified portfolio are, respectively, 0.608 to 0.814 percentage points higher than those of the US portfolio. The finding that international diversification reduces country-specific risk also applies to SRIs (Del Miralles-Quirós & Miralles-Quirós, 2017). Bauer et al. (2006) compare domestic and international SRI with conventional funds in Australia. They find that international SRI outperforms conventional funds, while domestic SRI underperforms conventional funds. The excess return of international SRI funds is 2.97% higher and that of domestic SRI funds is 3.22% lower than that of conventional funds. Kreander et al. (2005) also show that the Sharpe ratio and the Jensen measures of international ethical funds are higher than the FTSE World Index while international non-ethical funds have lower values than the FTSE World Index. Furthermore, domestic non-ethical funds have a slightly higher performance for risk-adjusted measures—such as Sharpe ratio, Treynor ratio, and Jensen's alpha—than domestic ethical funds.

In stark contrast to modern portfolio theory, agency theory (Aghion & Bolton, 1992; Grossman & Hart, 1992) argues that information is asymmetrically distributed between a principal such as a fund manager and an agent such as a company's CEO (Jensen & Meckling, 1976) and that investors typically have less information than corporate insiders such as the founder of a firm. As stated previously, screening for socially responsible assets increases search and monitoring costs. These costs can be reduced when investors are geographically close to an asset, because geographic closeness reduces the costs of screening (Cumming & Dai, 2010; Hornuf et al., 2022). Thus, local investments in SRI could ultimately generate better performance. However, as many companies and funds nowadays consider ESG criteria, an increasing universe of assets must be considered, with geographic proximity playing a less decisive role in ESG selection. Thus, we hypothesize:

H2: Studies that examine global SRI portfolios report better SRI performance than studies that examine regional SRI portfolios.

The impact factor provides information about the influence of a journal in the scientific community. Studies from the *Journal of Finance*, *Journal of Financial Economics*, and *Review of Financial Studies* are the most cited in finance scholarship (Alexander & Mabry, 1994; Calma, 2017). The impact factors of these journals are high, not only because the studies published in them are widely cited but also because they are the most prestigious and, thus, most attractive journals to publish in, not least because many of the Nobel laureates in economics have published their work in these outlets (Borokhovich et al., 2000; Borokhovich et al., 2011; Guo et al., 2016). The 50 most-cited finance studies are all published in the most prestigious journals (Arnold et al., 2003). Most citations in these finance journals are associated with a few known individuals, such as Eugene Fama, Stephen Ross, and Michael Jensen (Chung et al., 2001). In addition, many journals in which the 50 most-cited finance studies appeared are published by the University of Chicago Press (e.g., *Journal of Business*, *Journal of Political Economy*, *Journal of Law & Economics*), which supports the notion that top journals are often conservative and adhere more to modern portfolio theory than, for example, behavioral or social finance.

Many of the editors of the top finance journals have adhered to modern portfolio theory and, to a lesser degree, behavioral finance. For example, Michael Jensen, who in 1976 published “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure” in the *Journal of Financial Economics* also served as the first editor of the *Journal of Financial Economics*. His seminal article garnered more than 50,000 citations[4]. Robert Merton, who among others is well known for the Black–Scholes–Merton model, and Merton Miller, who is well known for the Modigliani–Miller theorem, were part of the committee of founders of the Society of Financial Studies, from which the *Review of Financial Studies* emerged. In 1972, Merton Miller, together with Eugene Fama, published the book *The Theory of Finance*. The first editor of the *Review of Financial Studies* was Michael Brennan, who was also an editor of the *Journal of Finance*. These are just a few of the renowned finance researchers who applied and developed modern portfolio theory and, at the same time, had a great influence on the most prestigious finance journals. Thus, it appears that finance journals are more conservative outlets.

[4] Semantic Scholar M. C. Jensen Citations 2021. Retrieved January 21, 2022, from https://www.semanticscholar.org/author/M.-C.-Jensen/48109845.

Moreover, academic institutions often pay scholars who publish in top journals higher salaries and additional financial bonuses (Fuyuno & Cyranoski, 2006). For example, Zhejiang Chinese Medical University pays its scholars $31,440 for a publication in *Nature* or *Science* (Jufang & Huiyun, 2011). Melbourne Business School pays its scholars $10,000 for a publication in the Top 50 list of the *Financial Times* (Macdonald & Kam, 2007). Therefore, in addition to recognition from academic peers, bonuses are another incentive to publish in top journals. Apart from the financial incentives, top-notch publications are an investment in the professional future, especially for assistant and associate professors (Oltheten et al., 2003; Swidler & Goldreyer, 1998). Maberly and Pierce (2007) investigate citations and self-citations of 94 journals that are referenced in the *Journal of Finance, Journal of Financial Economics*, and *Review of Financial Studies*. They note that the top three journals themselves are heavily cited and, together with other prestigious journals such as the *Journal of Financial and Quantitative Analysis* and *Journal of Business*, make up 43% of all citations examined. The studies in top-tier journals are mainly written by US scholars or scholars that work at US research institutions. Researchers that publish in top-tier journals often do not publish in lower-ranked journals, because such additional publications from lower-ranked journals negatively affect judgment of their achievements (Powdthavee et al., 2018).

As a result, finance scholars and journals are generally very conservative when it comes to introducing new theories and methods (Brooks et al., 2019). This fact is promoted, among other things, by members of editorial boards publishing frequently in their own journals (Hardin et al., 2008). Moreover, especially in economics, former doctoral students have significantly more publications in the *American Economic Review* and *Quarterly Journal of Economics*, if their supervisor is a member of the editorial board of the *Quarterly Journal of Economics* (Hilmer & Hilmer, 2011), which in turn makes the path dependency in the theories and methods more likely. Hardin et al. (2008) show that a publication in the *Journal of Finance* is a crucial criterion for becoming a member of the editorial board of financial journals. Moreover, Chan et al. (2015) argue that there is a coauthor network effect for the top three finance journals, such that editors favor weaker studies of former coauthors, which are likely to have a similar theoretical and methodological orientation. Therefore, the conservatism and high impact of the most prestigious journals are due not solely to their quality but also to authors' fixation on certain journals (Rubin & Rubin, 2021). Furthermore, scholars often strategically cite studies from the journals they wish to publish in, which promotes the theory bias of a particular journal and research strand. In summary, to publish in highly ranked finance journals, it is often not advisable to implement unorthodox methods. However, scholars should use state-of-the-art methods, collect recent data, and perform strict robustness checks to publish in finance journals (Cumming, 2020). To receive a revise and resubmit, scholars should generally follow the ideas of the classical modern portfolio theory, which implies that SRI should underperform relative to the market portfolio because of inferior diversification.

Many academic institutions today require their researchers to publish in top-notch financial journals to gain a higher reputation (Brooks et al., 2019). Graber et al. (2008) examine the number of publications by professors in Austria, Germany, and Switzerland to investigate whether the number and quality of publications have become more important over time.

Consequently, we conjecture that finance researchers who want to pursue an academic career must publish in highly ranked, peer-review journals. Moreover, Graber et al. show that researchers published the most in the most recent period they studied. The youngest scholars from 1997–2006 published 1.1 studies per year, the second-youngest scholars (1988–1996) only 0.7 studies, and older scholars even fewer. Moreover, quantity often comes at the expense of quality, and more publications often end up in lower-quality journals, which are also less conservative with regard to the methods applied than the most prestigious finance journals. However, if scholars publish well, it might be easier for them to build on the success and continue publishing in high-impact journals. The number of publications by one author might therefore indicate that a particular author follows the same theory. Thus, we hypothesize:

H3a: Studies published in journals with a higher impact factor are less likely to report outperformance of SRI.

H3b: Studies published in finance journals rather than other academic journals are less likely to report outperformance of SRI.

H3c: Authors who publish more frequently are less likely to report outperformance of SRI.

The capital asset pricing model (CAPM) developed by Sharpe (1964), Lintner (1965), and Mossin (1966) is still one of the most relevant capital market models. It describes the expected return of an asset with only one factor, namely risk, which is often referred to as beta. Beta results from the relationship of systematic risk with a financial asset. Since the 1970s, many empirical studies have added other factors that also have an impact on expected returns (e.g., Ross 1976). Most prominently, Fama and French (1993) introduced the three-factor model by adding firm size (SMB, small minus big) and the book-to-market ratio (HML, high minus low); Carhart (1997) later added the momentum factor (UMD, up minus down). These and other latent factors are often referred to as pricing anomalies, which the traditional CAPM cannot explain (Fama & French, 1996; Jarrow & Protter, 2016). The inclusion of these and other additional factors has been shown to reduce pricing errors and to improve the prediction of asset prices (Chen et al., 1986; Fama & French, 1993; Jagannathan & Wang, 1996). For example, the six-factor model introduced cash-based profitability (RMW, robust minus weak; instead of operating profitability of the five-factor model) and investment (CMA, conservative minus aggressive), which captures the difference in returns between firms with conservative and aggressive investment policies (Fama & French, 2015, 2018), while the q-model of Hou et al. (2015) contains size, investment, and return on equity, in addition to the CAPM. To investigate whether the underperformance or outperformance hypotheses hold in the factor model world, finance scholars often interpret alphas, the not explicitly modeled part of the factor model equation, as the performance difference between conventional and socially responsible assets.

Humphrey and Lee (2011) use one- and four-factor models to evaluate the performance of SRI and conventional investments in the Australian market. They show that SRI funds have higher abnormal rate of return than conventional funds when using Jensen’s (1968) alpha and lower alphas when using Carhart's (1997) alpha. The difference between SRI minus conventional alphas is positive for one-factor and negative for four-factor models, indicating declining SRI outperformance when including more factors. Thus, the simple one-factor model supports the outperformance hypothesis, which might be due to studies not accounting for all relevant factors correlating with SRI in the CAPM. When stock price, company size, and momentum are accounted for, modern portfolio theory trumps, and SRI no longer leads to outperformance. Gil-Bazo et al. (2010) find similar results. One-factor alphas of matched and unmatched SRI funds are positive, while four-factor alphas are negative. In summary, more factors in an asset pricing model explain excess returns better, and SRI outperformance consequently disappears. Thus, we hypothesize:

H4: Studies that consider more factors in a capital market model are less likely to report outperformance of SRI.

3 Data and Method

3.1 Systematic Literature Search and Meta-Analysis

We study the performance evaluation of SRI relative to a market portfolio using a large sample of studies from recent literature. Subsequently, we carry out a meta-analysis on the performance evaluation of SRI in these studies. To identify relevant studies, we conducted a search on IDEAS, which uses the RePEc (Research Papers in Economics) database, and ScienceDirect. We jointly searched for the terms "SRI," "return," and "performance" and also jointly for the terms "good performance," "financial performance," and "socially responsible investment." In a next step, we expanded the keyword search and took into account the keywords that were found in the studies we identified in the first step and added new keywords and keyword combinations on the topic of SRI. The keywords and keyword combinations are "ethical investment" and "return," "social, environmental, & ethical" and "return," "environmental, social, & governmental"/"ESG" and "return," "sustainable responsible investment" and "return," "sustainable investment" and "return," "responsible investment" and "return," "sin investment" and "return," "green investment" and "return," "Islamic investment" and "return," "corporate social investment" and "return," and "impact investing" and "return." Last, we conducted a Google Scholar search, but with slightly different keywords, because our previous keywords resulted in far too many hits. We searched for "ESG fund" or "ESG stock" or "ESG bond" and, respectively, "excess return" as well as "SRI fund" or "SRI stock" or "SRI bond" and, respectively, "excess return." We chose these keywords because, in essence, we are interested in the outperformance of SRI and ESG funds, stocks, and bonds. Studies in this more concise search, however, often also included the set of keywords from our original search in IDEAS and ScienceDirect.

We conducted the systematic literature search from December 2019 to April 2021. Our search resulted in 5,845 hits from all three databases. First, we checked the sample for duplicates. If a study had first appeared as a working paper and was later published as a journal article, we only considered the final journal article; otherwise, we considered the working paper. When the title of a study changed, we then identified duplicates using the authors' names. This led to the exclusion of 1,507 studies. Second, to be included in the meta-analysis, a study had to fulfill three additional inclusion criteria:

1. Empirical study: The measurement of SRI performance is at the core of our meta-regression. Therefore, we excluded 2,505 theoretical and narrative studies as well as literature overviews and previous meta-analyses.
2. Academic study: We considered published and unpublished studies, but excluded 114 student research papers. Student research papers are bachelor and master theses, but not doctoral theses.
3. Language of study: The studies had to be written in English. We excluded 36 studies that were written in other languages.

Furthermore, for our analyses, studies had to contain performance measures of SRI. We excluded 1,530 studies that did not report such performance measures. Islamic investment is often closely related to SRI (Renneboog et al., 2008), because ESG-based investment principles also originated from core religious values (Kiymaz, 2012; Schwartz, 2003). We included Islamic investment studies in our sample if either Islamic investment had been categorized as SRI by the authors of the respective study or Islamic investment was compared with SRI.[5]

These exclusion criteria left us with 153 studies and 1,047 observations of SRI performance. Studies were coded multiple times if they included SRI performance measures of multiple disjunct regions, had more than one sample period, or used more than one method for performance evaluation. Compared with other meta-analyses, ours is to the best of our knowledge the most recent and most sophisticated analysis and is also based on the largest sample. Rathner's (2013) meta-analysis includes 25 studies, and Revelli and Viviani's (2015) is based on 120 studies. The meta-analysis of Friede et al. (2015) contains 60 studies, which is the sum of 35 studies with significant positive, negative, and non-significant results and 25 other meta-analyses. Ait El Mekki (2020) investigates 103 studies in his meta-analysis, and Horváthová (2010) considers 37 studies. In addition to these studies, we further report results of 13 event studies and studies using peculiar methods, which has not been not done by other authors and allows for a more comprehensive picture of SRI performance evaluations.

Overall, our sample includes 124 studies and 694 observations that do compare SRI with the market portfolio as well as 62 studies and 353 observations that do not consider a benchmark. Overall, 33 of these studies allow for both: a comparison with a market portfolio and no comparison. To be considered for the meta-regression, studies must have compared SRI performance with a market portfolio or conventional portfolio. Alternatively, the SRI performance comparison could be made with an index. The baseline meta-regression is based on funds and stocks and consists of 119 studies and 666 observations, in which SRI performance measures are reported and all explanatory variables are available. SRI bonds are excluded in the meta-regression because fixed income returns are different in nature and the sample size is insufficient to be separately analyzed in a meta-regression (3 studies and 26 observations).

[5] We then consider SRI performance for the analysis "SRI Without Market Comparison."

While SRI as a concept has been analyzed in the literature for more than two decades (e.g., Mueller, 1991; Caplan et al., 2013), studies on ESG investing have only recently gained prominence. This fact is also reflected in our sample. Of 119 studies in our sample, 111 studies examine SRI and only 6 studies ESG. The two concepts are examined together in another 2 studies (see Figure 1). Due to the small number of ESG papers, we do not differentiate between SRI and ESG in our meta-analyses, but use the global term socially responsible investments. Figure 2 reports the analysis of the two concepts over time in relation to all publications respectively for five year periods.

[Figures 1 and 2 about here]

3.2 Variables

3.2.1 Dependent Variable

We construct two dependent variables to evaluate the SRI performance as reported in the respective study: *performance differences* of SRI relative to the market and *outperformance of SRI*. First, the SRI literature is heterogeneous with regard to how the performance of SRI is evaluated and with which method it is empirically studied. For that reason, we needed to make the various measures of performance evaluation comparable across studies, particularly as they relate to different asset classes such as funds, stocks, and bonds. Reporting simple differences of SRI and the market portfolio would not be expedient, because, for example, differences in Jensen's alpha cannot be meaningfully compared with the differences in the Sharpe ratio. To standardize the differences in performance, we calculate the percentage performance differences between the annual performance of SRI and the market portfolio, which also has an intuitive interpretation. This allows us to investigate not only the outperformance of SRI but also the *extent* of the outperformance of SRI. Second, and in line with previous meta-analyses, we generate the dummy variable *outperformance of SRI*, which is equal to one if SRI outperforms the market portfolio in the respective study and zero otherwise.

3.2.2 Explanatory Variables

To investigate H1, we compare the annual mean performance measures of SRI with the respective market portfolio. Because studies employ many different performance evaluation methods, we group them into seven categories. First, we compare plain annual returns of SRI with the market portfolio. Second, we compare the excess returns of SRI (i.e., the difference between SRI and the risk-free rate) with the excess return of the market portfolio. Third, we compare risk-adjusted measures of SRI with risk-adjusted measures of the market portfolio, such as Sharpe ratio (SR), Treynor ratio (TR), information ratio (IR), Modigliani Modigliani ratio (M^2), Sortino ratio (SoR), and omega ratio (OR).[6] Fourth, we compare other SRI returns

[6] SR considers excess return over its standard deviation, and TR measures excess return over systematic risk. IR includes abnormal return over the unsystematic risk. Furthermore, M^2 extends SR by multiplying it by the standard deviation of the market index. SoR considers excess return over the standard deviation of negative returns. Last, OR weights the probability of gains and losses of the minimum acceptable return.

that were retrieved from more complex models with market portfolio returns. These models are, for example, four-factor adjusted abnormal returns, excess standard deviation–adjusted returns, or data envelopment analysis (DEA) outputs. Fifth, we compare SRI alphas of factor models and SRI market-timing models with the alphas of the market portfolio. Table 2 summarizes alphas from various types of factor models under the term "factor models." The category efficiency measures include the efficiency scores of DEA and free disposal hull (FDH). DEA uses input and output variables to measure fund performance and evaluate the relative efficiency of decision-making units (Basso & Funari, 2001). FDH is a non-convex version of DEA (Abdelsalam et al., 2014). Seventh, some studies also include an SRI dummy variable in a regression model as a performance measure. The Appendix provides a detailed explanation of the methods.

To test H2, we investigate the outperformance of global SRI portfolios as compared with regional SRI portfolios and consider the following regions from which studies collected and examined data: *Asia*, *Europe*, *global*, *rest of the world*, and the *US.* We generate five dummy variables for these regions, which take the value one if the data for the respective observation comes from that region and zero otherwise. To test H3a, we consider the variable *impact factor*, which is the value of the 2019 impact factor of the journal in which the respective study was published. To investigate H3b, we include a dummy variable *finance journal*, which is equal to one if an observation is published in a finance journal and zero otherwise[7]. To test H3c, and in line with Rathner's (2013) meta-analysis, we consider the variable *number of publications*, counting the number of publications that one author has contributed to our dataset. The number of publications by one author captures the possibility that a particular author consistently follows the same theory and the respective outcomes cluster.

Finally, prior research has shown that in addition to market risk, many other factors explain returns (Carhart, 1997; Fama & French, 1996, 2015). While ESG might constitute an additional factor in itself (Pedersen et al., 2021), it might well correlate with exposures to other previously identified factors. If that is true, controlling for more factors should decrease the probability that a study finds SRI outperformance. To test H4, we consider different dummy variables that capture the performance measurement methods. The categories are the same as when testing H1. Moreover, to identify H4 more thoroughly, we split alpha into *one-factor model Jensen's alpha* and *multi-factor model alpha*. In the meta-regression, the categories for performance measures are thus *mean returns*, *excess returns*, *risk-adjusted measures*, *other returns*, *one-factor model Jensen's alpha*, *multi-factor model alpha*, *efficiency measures*, and *SRI dummy variables*.

3.2.3 Control Variables

We included several control variables in our meta-regression, following prior meta-analyses by Rathner (2013) and Tully and Winer (2014). First, we include four dummy variables to control for the different sample periods of the data on which the respective study is based: *SPeriod*

[7] SJR Finance Journals. Retrieved 2021, from https://www.scimagojr.com/.

1981–1990, *SPeriod 1991–2000*, *SPeriod 2001–2010*, and *SPeriod 2011–2020*. These variables are equal to one if the sample period in the study falls in one of the predefined periods and are zero otherwise. If the sample in a study covers more than one period, only the dummy variable with the larger share of the sample period in a study is equal to one. Given the recent outperformance of SRI (Solberg, 2022), we conjecture that more recent periods are more likely to show SRI outperformance in the meta-regression. This conjecture is generally in line with Rathner's (2013) finding that SRI performance differs for various subperiods. Moreover, López-Arceiz et al. (2018) and Chung et al. (2012) both investigate US fund data and calculate four-factor alphas. For the period 2004–2014, López-Arceiz et al. (2018) report a positive alpha for SRI only and an outperformance of SRI in terms of alpha when compared with conventional investments. For the period 2000–2009, Chung et al. (2012) report a negative alpha and an underperformance of SRI in terms of alpha when compared with conventional investments. Thus, different sample periods affect SRI performance. Finally, including dummy variables for the respective sample period is important because SRI has gained tremendous practical relevance in the past decade and gained increasing interest from the academic community. In support, the period 2001–2010 entails more observation in our sample than the period 1981–1990 (314 vs. 19 observations).

Because risk and other factors can affect portfolio performance, we consider different empirical measures and methods to control for their impact in the meta-regression. Traditionally, research has treated alphas and betas as fixed in the standard CAPM and other standard factor models, such as the Fama–French model. More recently, so-called conditional factor models (Ang & Kristensen, 2012; Gagliardini et al., 2020) also allow for time-varying alphas and betas. Unconditional factor models may be biased, because trade and many other macro-economic factors change over time (Chen & Knez 1996; Ferson & Schadt, 1996). To account for the difference, we include the variable *conditional evaluation*. This dummy variable is equal to one if a conditional factor model is used in a study and zero otherwise. As Rathner (2013) notes, conditional and unconditional performance evaluations lead to different results even within a single study. For example, Bauer et al.'s (2006) study shows that SRI underperforms the market portfolio when relying on a four-factor alpha model and outperforms it using a conditional four-factor alpha model.

In line with Rathner (2013), we include a dummy variable that captures whether a matching procedure was used for the performance evaluation of SRI in the respective study. Some studies match the SRI funds in their dataset to the same number of conventional funds (*one-to-one matching*) (Gregory et al., 1997; Humphrey & Lee, 2011; Kreander et al., 2005), others match the SRI funds to more conventional funds (*one-to-many matching*) (Bauer et al., 2005; Bollen, 2007; Hamilton et al., 1993), and still others do not use a matching procedure at all. Matching helps researchers compare apples with apples in an empirical study. If assets are not matched, the differential performance of SRI might simply be due to the different underlying factors of an asset that are correlated with ESG criteria. We also consider the number of underlying matching criteria in a study by including the variable *NumMatchCriteria.*

We further control for the type of market portfolio that was used as a benchmark in the respective study. Some studies use a sample of conventional assets, such as funds, stocks, or bonds, and compare it with the same assets that, however, score high on ESG criteria. Sometimes these studies apply matching techniques to make the comparison more valid. Others studies do not compare conventional assets with assets scoring high on ESG criteria, but rather define the benchmark as market portfolio and consequently compare SRI performance with, for example, the MSCI World Index. To account for the different study types, we included two dummy variables that categorize the respective benchmark. The dummy variable *same asset class benchmark* is equal to one if the benchmark consists of individual assets from the same asset class, such as funds, stocks, and bonds, and is zero otherwise. The dummy variable *index benchmark* is equal to one if the respective study uses a particular national or international index as benchmark and is zero otherwise.

When analyzing the performance of an asset, the failure to take into account failed funds or stocks can lead to a survivorship bias and, thus, to incorrect performance evaluations (Brown et al., 1992). However, not all studies consider survivorship bias (Rathner, 2013). Therefore, we also control for the variable *survivorship bias*, which is equal to one if the respective study considers failed assets and zero otherwise.

Another quality criterion for a publication is whether a study was published in a peer-reviewed journal. While peer-reviewed publications might signal higher-quality standards, not considering studies that have been published in outlets without a peer-review process might lead to a publication bias (Tully & Winer, 2014). We therefore use the dummy variable *peer-reviewed*, which is equal to one if the respective study was published in a peer-reviewed journal and zero otherwise. Furthermore, we include a dummy variable to differentiate between stocks and funds in the meta-regression for equity. The control variable *Stock* takes the value one if a study reports the performance of SRI stocks and 0 if a study reports the performance of SRI funds. Furthermore, we control for databases used in the respective study. We therefore consider dummy variables capturing the four most commonly used databases for financial data in our sample—*Bloomberg*, *CRS*, *Morningstar*, and *Thomson Reuter*—and use *other* for the remaining databases. The dummy variables are equal to one if the particular database is used in a study and zero otherwise. ESG taxonomies can differ significantly from each other (Dumrose et al., 2022). The *ESG taxonomy* control variable takes on a value of one when a study reports a source for an ESG, SRI, or ethical taxonomy provided by a professional data provider. In addition, we also control the precise ESG taxonomies used in the respective study. We again consider dummy variables for the commonly used taxonomies in our sample—*Bloomberg*, *Morningstar* (incl. Sustainalytics), *SIF* (Sustainable Investments Forums, incl. Eurosif and US SIF), *SocialFunds.com*, and *Vigeo*—and *other* for the remaining taxonomies.

3.3 Empirical Models

Because our first dependent variable is a continuous variable, we employ ordinary least squares (OLS) regressions. We specify the following baseline regression model:

$$
\begin{aligned}
Performance\ difference_{ij} &= \beta_0 + \beta_1 \mathbf{Region}_{ij} + \beta_2 \text{Impact factor}_j + \beta_3 \text{Finance journal}_j \\
&+ \beta_4 \text{Number of publications}_j + \beta_5 \mathbf{Method}_i + \beta_6 \mathbf{Controls} + \varepsilon_{ij},
\end{aligned}
$$

where *performance difference* refers to percentage differences between the annual SRI and the market portfolio performance that was reported in observation *i* in study *j* and **Region** is a vector of the dummy variables capturing *Asia*, *Europe*, *rest of the world*, and *US.* We excluded the baseline category *global* in the regressions. *Impact factor* is the 2019 journal impact factor of the journal in which the study appeared, *finance journal* is a dummy variable indicating whether the study was published in a finance journal or not, *number of publications* indicates the number of studies by a specific author in our dataset, and **Method** is a vector for the performance measure dummies *excess returns*, *risk-adjusted measures*, *other returns*, *one-factor model Jensen's alpha*, *multi-factor model alpha*, and *efficiency measures*. *SRI dummy variables* is the baseline category and is excluded in the regressions. **Controls** is a vector of the control variables as outlined in section 3.2.3, and ε is the error term.

In an alternative model, we analyze whether an observation reports SRI outperformance or not. We consequently estimate a probit regression. The specifications are identical to the OLS regression. The probit regression model takes the following form:

$$
\begin{aligned}
Prob\left[Outperformance\ of\ SRI_{ij}\right] &= \Phi\big(\beta_0 + \beta_1 \mathbf{Region}_i + \beta_2 \text{Impact factor}_j + \beta_3 \text{Finance journal}_j \\
&+ \beta_4 \text{ Number of publications}_j + \beta_5 \mathbf{Method}_i + \beta_6 \mathbf{Controls} + \varepsilon_{ij}\big).
\end{aligned}
$$

4 Results

4.1 Performance of Funds, Stocks, and Bonds Without a Market Comparison

Some studies report SRI returns by comparing them with another "ethical" investment, such as Shariah-compliant portfolios, or do not compare SRI performance with a benchmark at all. As a starting point, we therefore report the plain performance of SRI funds, stocks, and bonds in Table 1. In particular, we calculate the average annual value of the different performance measures that have been reported in the respective studies, which are annual returns, excess returns, risk-adjusted measures, other returns,[8] factor models, and efficiency measures. We find that the financial performance of SRI varies substantially and depends on the respective asset class and performance measure.

[8] Other returns for funds without a market comparison are, for example, DEA outputs.

The plain annual return for SRI funds is 3.8% (21 studies; 47 observations[9]). Based on a significantly smaller number of observations (2 studies; 8 observations), studies investigating individual SRI stocks report an annualized return of 23.2%. Two studies with 23 observations contain mean returns for SRI bonds, which yields a mean return of 7.0%. SRI funds are reported to result in mean excess returns of –0.4% (2 studies; 4 observations), while SRI stocks are reported to result in higher mean excess returns of 1.2% (4 studies; 18 observations). The bond i-spread can be interpreted as the excess return of a bond and was reported to be 47.5 bps in one study. For SRI funds and stocks, some studies also consider risk-adjusted measures without a benchmark comparison. The average annual value of risk-adjusted measures for funds is 0.078 (9 studies; 22 observations). The mean risk-adjusted measure for SRI stocks is 0.824 per year (3 studies; 19 observations). By comparison, the three-year Sharpe ratio of the MSCI World Index (2022) is 0.880. Other returns for funds (2 studies, 2 observations) are 0.7%, which constitute DEA returns (output) in both cases. Most studies use factor models to assess SRI performance. The mean alpha for SRI funds is –0.322 (30 studies; 139 observations) and –0.014 for SRI stocks (12 studies; 59 observations), which suggests that most SRI portfolios presumably underperform the market. One study used a factor model to evaluate bond performance and reported a mean alpha of 0.006 (8 observations). Only a few studies calculate efficiency measures[10] for funds. The average value for the efficiency measures is 106.7% (3 studies; 3 observations). An efficiency score of 100% implies that the fund is fully efficient, and a higher score indicates inefficiency.

[Table 1 about here]

4.2 SRI with a Conventional Benchmark Comparison

To test H1, which states that SRI underperforms the market portfolio, we analyze the reported annual average performance of SRI funds, stocks, and bonds relative to the reported performance of the market portfolio. Table 2 shows the results. Panels A–Q differentiate the performance measures that have been calculated in the various studies as outlined in section 3.2.2. Line (1) of the panels shows the average performance of SRI for each category of performance measure, and line (2) shows the average performance of the market portfolio. First, we compare the average performance measures of SRI with the market portfolio and conduct a Wilcoxon rank-sum test to check whether the reported performances are on average statistically different from one another ((1)–(2)). Although returns are often not normally distributed (Harris & Kucukozmen, 2001), we nevertheless conduct a two-sample t-test as a robustness check.[11] Second, in line (3) we test whether *performance difference* of SRI, the dependent variable in

[9] We included multiple observations for a study when the study used multiple independent regions, independent sampling periods, or more than one method of performance evaluation.
[10] These are efficiency scores and not returns calculated as DEA outputs.
[11] The results of a two-sample t-test largely confirm the results of the Wilcoxon test. When the results of the t-test differ from those of the Wilcoxon test, they are either not normally distributed according to a Shapiro–Wilk test or have small sample sizes, which indicates that a non-parametric test is the better choice. The results of the t-tests are available on request.

our meta-regression, is statistically different from zero using a sign test.[12] Third, the mean of *outperformance of SRI* in line (4), our alternative dependent variable in the meta-regression, is the share of observations that report SRI outperformance relative to no outperformance. We conduct a Wilcoxon signed-rank test to determine whether significantly more than 50% of the observations in our dataset report SRI outperformance.

To interpret Table 2, consider the example of plain annual returns in Panel A. In lines (1) and (2), we report that the average annual return of SRI funds is 6.1% and the average annual return of the market portfolio is 6.3%. Thus, the difference is 0.2 percentage points, which shows that SRI underperforms, which is in line with traditional finance theory. To compare the returns of SRI funds with the market portfolio (line (1) vs. line (2)), we first conduct a Wilcoxon rank-sum test, which shows that the performance of the SRI funds is not significantly lower than the performance of the market portfolio (p = 0.600). Second, to make the various performance measures comparable, we calculate the percentage performance difference between SRI funds' performance and the market portfolio performance for each observation in our dataset and report the mean value. This is the dependent variable in our OLS regression. In the case of plain annual returns, the mean performance difference is 51.5%, which indicates that there are positive outliers in the dataset. Note that this figure does not result from the percentage difference of line (1) and line (2), but is the mean value of the percentage difference of each observation in our dataset. For example, Humphrey and Lee (2011) report an annual return of 8.7% for SRI funds and 8.6% for the market portfolio. The difference is therefore 0.1 percentage points, and the percentage difference between the two returns is 1.5%.[13] However, some studies report more extreme values, such as an annual return of 5.5% for SRI funds and 0.1% for the market portfolio (Hwang et al., 2011). In this case, the difference is 5.4 percentage points, and the precise percentage difference is 6000%. Because the percentage differences are not normally distributed, we test whether *performance difference* is statistically different from zero using a sign test. We find that the percentage difference in the performance is weakly significant at the 10% level relative to zero (p = 0.057). In line (4), we count the observations in our dataset that report SRI outperformance and find that 39.1% reported SRI outperformance while 60.9% consequently did not. Finally, we test whether the number of observations showing SRI outperforms is beyond 50% using a Wilcoxon signed-rank test. The result of 39.1% is statistically different from 50%, which indicates that most observations find no SRI outperformance.

Overall, we find that the performance of SRI funds is different from the market portfolio for three out of seven performance measures at conventional statistical levels. On average, excess returns (Panel B), factor models (Panel E), and the dummy variables (Panel G) show significant differences between SRI and the market portfolio. While excess returns and dummy variables show negative average differences ((1)–(2)) of respectively –1.8 and –15.8 percentage points, the difference for factor models is positive at 1.8 percentage points. The *performance difference*

[12] Note that the value of *performance difference* in line (3) is not the percentage difference of line (1) and (2) but the average of individual percentage differences for each observation in our dataset.

[13] The percentage difference results from the precise values and is calculated as $\frac{8.731\%-8.602\%}{8.602\%} = 0.01499 = 1.5\%$.

of excess returns in line (3) of Panel B is also statistically different from zero and indicates that SRI underperforms the market portfolio. Moreover, the number of observations reporting SRI outperforms is below 50% for most performance measures, except risk-adjusted measures (Panel C).

For stocks, we find a positive difference between SRI and the market portfolio in terms of excess returns (Panel I) of 7.4 percentage points annually. The difference is statistically significant at the 5% level. Furthermore, annual returns (Panel H), excess returns (Panel I), and risk-adjusted measures (Panel J) show SRI performance differences that are significantly different from zero, all indicating SRI outperformance. The number of observations in the dataset showing SRI outperforms for stocks is above 50% for annual returns, excess returns, and risk-adjusted measures, whereas factor models (Panel K) more frequently report no outperformance. These results are statistically significant at conventional levels. In case of bonds, the *performance difference* of factor models (Panel P) is positive and significantly different from zero. The number of observations in the dataset that shows SRI outperforms for bonds is above 50% for annual returns (Panel M) and factor models and statistically significant at conventional levels.

[Table 2 about here]

In summary, the performance comparison in Table 2 shows mixed results for H1 that SRI underperforms the market portfolio. We do not find a robust tendency for SRI performance under- or outperformance for either certain asset classes or performance measures. If anything, our results lend support to the conjecture that, on average, there are no differences between the performance of SRI and the market portfolio.

4.3 Summary Statistics

Table 3 and Table A1 in the Appendix report summary statistics and a correlation matrix for the variables in the meta-regression. The correlation coefficients are generally small. We also consider variance inflation factors (VIF) in the regressions and find no evidence of strong multicollinearity. To make the performance measures from different studies comparable, we consider *performance difference* and the dummy variable *outperformance of SRI*. The average performance difference of SRI relative to the market portfolio is 1.5%, which indicates that unconditionally the financial performance of SRI measured by various performance measures is on average 1.5% higher than that of the market portfolio. The average of *outperformance of SRI* is 0.453, which means that 45.3% of the observation in our sample report SRI outperformance, while the remaining observations do not find a difference between SRI and the market portfolio or SRI underperformance.

The studies in our dataset have an average impact factor 1.6, and an author published on average 2.4 studies in our sample. With 37.1% and 36.6%, the majority of the observations represent assets from Europe and the US, respectively, followed by Asia and the rest of the world. Not even 3% of the observations represent genuine global assets, but some regional portfolio. Almost one-third of the observations are published in finance journals.[14] Three-quarters of the observations are published in peer-reviewed journals, while the remaining observations come from journals without a peer-review process, books, or publications that are still in the working-paper stage. Performance evaluation indicates the method used to identify the performance difference between SRI and the market portfolio in the respective study and whether a conditional model was used. Most often, scholars reported multi-factor model alphas, mean returns, or some other type of risk-adjusted measures. With 47.2%, most observations rely on the sample period 2001–2010, followed by the period 1991–2000 accounting for 30.3% of the observations. Scholars use data from the most recent sample period 2011–2020 in 21.6% of the observations. In sum, 6.5% of the observations rely on one-to-one matching and 17.3% one-to-many matching. Over all observations, studies have on average applied 0.6 matching criteria. Three-quarters of the observations investigate funds, and the remaining quarter analyzes stocks. With 79.9%, the majority of observations compare SRI funds or stocks with conventional funds or stocks, while 20.1% of the observations take an index as a benchmark. Almost half the observations consider a potential survivorship bias, while slightly more than half do not. Most often, the observations in our sample rely on financial data from Thomson Reuters, followed by CRSP and Morningstar. Overall, 60.5% of the observations report which taxonomy was used for ESG. Most of the observations in our sample use Morningstar, followed by SIF and Vigeo.

[Table 3 about here]

4.4 Publication Bias

When journals prefer to publish studies that only provide evidence of SRI underperformance or significant statistical results, a publication bias can occur in our meta-analysis. A common method to identify the presence of a potential publication or selection bias in a meta-study is funnel plots (Harbord et al., 2006). The usual representation of funnel plots is a scatterplot in a Cartesian coordinate system, with effect sizes—in our case the *performance difference*—plotted on the x-axis and sample sizes or standard errors of the respective studies on the y-axis.[15] A symmetrical shape should result in the funnel plot if the studies included in the meta-analysis are balanced in their results considering natural statistical variability. Larger studies should produce more concise results that are closer to the mean of all study results. If the scatterplot is asymmetrical on one side of the mean due to many small studies, this would suggest unpublished or unconducted studies with conflicting results and thus a publication bias.

[14] Overall, 40 studies in our sample were published in finance journals. If each study is considered only once, the average impact factor is 1.7, and an author in our sample published an average of 2.3 studies.
[15] We chose sample sizes because few studies in our sample report standard deviations or variances. However, this does not diminish the validity of the funnel plot.

Figure 3 shows a funnel plot for the studies investigating SRI funds. Not all studies in our sample report the sample size of their empirical analysis, leaving us with 418 observations. Furthermore, we deleted 5 extreme outliers to increase the readability of the figure, so the funnel plot ultimately consists of 413 observations. The resulting symmetrical shape of Figure 3 indicates that there is no publication bias in studies analyzing SRI funds. Figure 4 shows the funnel plot for SRI stocks. Unfortunately, only 28 of 151 observations reported sample sizes for this subsample.[16] However, there is no apparent asymmetry around the mean performance difference for stocks either.

[Figures 3 and 4 about here]

4.5 Meta-Regression

To test H2, which states that studies that examine global SRI portfolios report better SRI performance than studies that examine regional SRI portfolios, we run meta-regressions and include the dummy variables *Asia*, *Europe*, *rest of the world*, and the *US*; *global* constitutes the baseline category. In the baseline meta-regression, we consider all observations except those measuring bond performance, because their performance is not comparable with that of funds and stocks. Table 4 reports the main results. As a robustness check, we consider subsamples of funds, stocks, and bonds in Tables 5–7. In column 1 of each table, we estimate an OLS model with *performance difference* as the dependent variable; in column 2, we estimate a probit model with *outperformance of SRI* as the dependent variable.[17] For the probit regressions, we report average marginal effects. The numbers of observations for stocks are relatively small and even smaller for bonds, which is why these results need to be interpreted with caution. To give a complete picture, we report an OLS regression for the bonds-only subsample as well.[18]

In the baseline meta-regression in Table 4, in which we jointly consider funds and stocks, we find a significant result for *US*. The results of the OLS model indicate that when investing in portfolios in the US only, the performance of SRI relative to the market portfolio decreases by 22.5 percentage points compared with global portfolios, which is significant at the 5% level. The probit model reports that the probability of *outperformance of SRI* decreases by 11.3% if US portfolios are compared to global portfolios, but this is only weakly statistically significant.[19] When considering the subsample of funds in Table 5, we find that investing in

[16] We take comfort in the fact that there is no apparent theoretical reason why reporting the sample size should eliminate a potential publication bias. If anything, one would expect that the comparatively small sample of 28 observations leads to artificially skewed results in the funnel plot. However, the funnel plot for this small sample shows no publication bias, which should become even more evident in a larger sample.

[17] The descriptive statistics for the dependent variables are available in Table 3.

[18] However, we do not fit a probit model for bonds, because probit might be biased for small samples (26 observations) (Stone & Rasp, 1991) and too many explanatory variables predict *outperformance of SRI* perfectly. The OLS model for bonds does not contain all variables, because the studies on bonds in our dataset do not take all explanatory and control variables into account. For example, performance is calculated using only risk-adjusted measures, factor models, and SRI dummy variables. Nevertheless, the variables we included in Table 7 explain 32.8% of the variation in *outperformance of SRI* for bonds.

[19] The percentage points figures correspond to the difference in the percentage performance differences between the annual performance of SRI and the market portfolio, which we calculated to standardize study results. They do not correspond to plain return differentials of SRI and the market portfolio.

rest-of-the-world portfolios reduces SRI performance relative to the market portfolio by 26.6 percentage points compared to investing in global portfolios. However, the coefficient is only weakly significant. The probit model reports a 20.0% lower probability to find SRI outperformance for rest-of-the-world portfolios compared to global portfolios, which is significant at the 5% level. Next, we consider the subsample of stocks in Table 6. We find that investing only in Europe or the rest of the world reduces the performance of SRIs relative to the market portfolio when compared with global investing**.** Moreover, the probability of finding studies that report outperformance of SRI is significantly lower for observations that examine regional portfolios from Europe compared with global portfolios. Regarding bonds in Table 7, we only consider *Europe* and *US* given data limitations and find that the relative performance of SRI is worse for US samples than European samples. Investing in bond portfolios in the US decreases the performance of SRI relative to the market by 57.0 percentage points compared with investing in European bond portfolios, which is significant at the 5% level.

In summary, we find evidence for H2 that studies that examine global SRI portfolios report better SRI performance than studies that examine regional SRI portfolios. These findings are comparable to those of other studies, showing that international diversification increases expected returns for a given level of risk (e.g., Goetzmann & Ukhov, 2006), which is in line with modern portfolio theory.

To test H3, we use *impact factor*, *finance journal*, and *number of publications* as variables of interest in our meta-regressions. Table 4 reports the results. We find that the performance of SRI decreases when the *impact factor* increases. More precisely, if the journal impact factor increases by one, the performance of SRI relative to the market decreases by 9.4 percentage points. The probit model reports a corresponding 4.5% lower probability to find SRI outperformance if the impact factor of a journal increases by one. Both coefficients of *impact factor* are statistically significant at the 1% level. If findings were reported in a finance journal, the performance of SRI relative to the market portfolio decreases by 34.8 percentage points. The probit model reports a 22.6% lower probability to find SRI outperformance if the results are published in a finance journal. Both coefficients are significant at the 1% level. Furthermore, we find that authors who publish more frequently are less likely to report better performance of SRI. If the number of publications of an author increases by one, the performance of SRI relative to the market decreases by 4.7 percentage points; the probit model reports a corresponding 3.4% lower probability to find SRI outperformance, respectively significant at the 1% and 5% level.[20]

Our findings remain robust for the subsample of only funds in Table 5. The coefficients of *impact factor*, *finance journal*, and *number of publications* in the OLS and probit regressions

[20] We are also interested in whether researchers can increase their chances of publishing in peer-reviewed journals by writing more articles. Therefore, we test whether more publications by scholars have an impact on the quality of their output—more precisely, whether frequent publishing affects the publication in peer-reviewed journals. The number of publications by authors ranges from 1 to 7 in our dataset, with a median of 2. We find that authors who publish at least twice have a 79.9% chance of publishing in a peer reviewed-journal, while the probability for authors with fewer than two publications is only 72.4%. A two-sample t-test shows that this difference is significant at the 5% level. Thus, we find some form of path dependence, in that authors who publish more frequently will eventually publish in peer-reviewed journals.

are all negative and remain significant. However, the results are no longer significant for the stocks and bonds subsample, which is most likely due to the significantly reduced sample size.

Overall, we find support for H3a, H3b, and H3c; that is, a higher impact factor, publishing in finance journals, and publishing more frequently decreases the probability of SRI outperformance. These results are consistent with previous literature, which argues that modern portfolio theory is difficult to displace and that traditional finance theory is better followed when scholars want to publish in the top journals and in particular top finance journals (Chung et al., 2001; Cumming, 2020). Moreover, scholars who publish more and are frequently cited in the community are often full professors at top universities (Amara et al., 2015). Apparently, these researchers know the publication game, are more likely to adhere to traditional finance theory, and publish more often.

Finally, we investigate H4 and test whether considering more factors in a model decreases the outperformance of SRI. We consider different performance measures as variables of interest, in particular *excess returns*, *risk-adjusted measures*, *other returns*, *one-factor model Jensen's alpha*, *multi-factor model alpha*, *efficiency measures*, and *SRI dummy variables* with *mean returns* as the baseline category. The results of the OLS model indicate that using *multi-factor model alpha* as the performance measure decreases the *performance difference* of SRI relative to the market by 22.2 percentage points compared with using *mean returns*. The coefficient is significant at the 5% level. The coefficient for *multi-factor model alpha* in the probit model is also negative but not statistically significant at conventional levels. Our results provide evidence that SRI outperformance decreases when multi-factor models are used. We show that studies that take more factors into account when assessing performance report lower SRI performance on average, in support of H4.

We are further interested in whether the particular performance measure used in a study published in a journal with a high impact factor or in a finance journal makes a difference. Therefore, we examine whether publications in journals with high impact factors or publications in finance journals have an effect on the choice of performance measure used in a study. Table 8 reports the results. First, we estimate an OLS model with *impact factor* as the dependent variable. Second, we perform a probit model with *finance journal* as the dependent variable. The explanatory variables are the performance measures outlined in 3.2.2 with the exception of the mean returns; *efficiency measures* constitutes the baseline category. The results show that *excess returns* is the only performance measure that is positively correlated with a journal's impact factor. Using the one-factor model performance measure *excess returns* increases the impact factor by 1.2 compared with using *efficiency measures* as a performance measure. In the probit model, *excess returns*, *risk-adjusted measures*, *other returns*, *one-factor model Jensen's alpha*, *multi-factor model alpha*, and *SRI dummy variables* all have positive coefficients. The coefficients of *one-factor model Jensen's alpha*, *multi-factor model alpha*, and *SRI dummy variables* are significant at the 1% level, indicating that these methods increase the probability of a study being published in a finance journal compared with using *efficiency measures* as a performance measure.[21]

Some control variables we consider in the meta-regression also reveal notable results. Because of the increasing demand for SRI in recent years, we would have expected better performance of SRI when studies consider more recent data. Because of the low number of observations for the period 1981–1990, we combined the sample periods 1981–2000. We find that SRI performs better than the market for the oldest sample period (*SPeriod 1981–2000*) than the latest sample period (*SPeriod 2011–2020*). However, as Table 4 shows, the performance of SRI relative to the market increases by 42.5 percentage points for the period 1981–2000. The coefficient for *SPeriod 1981–2000* in the probit model is also positive but only marginally significant. However, the better performance of SRI than the market portfolio disappears after that period. One potential reason for this finding is that investors became aware of the outperformance and the "anomaly" disappeared over time because markets became more efficient. Another explanation might be that top journals do not publish replications, and the outperformance of SRI had already been shown. For example, Alm and Reed (2015) find that only 2.4% of 333 economics journals indexed in Web of Science explicitly mentioned on their websites that they publish replication studies. The reasons behind journals not publishing replications are lack of space and editors' perception that replications cannot compete with the original research. New results are therefore given priority in the publication process. Recognizing this behavior, researchers avoid replications, knowing that few peer-reviewed journals publish replications and most prefer novel results (Alm & Reed, 2015). The first studies reporting SRI outperformance might have been published because SRI outperformance was a surprising result. Finally, SRI outperformance simply might not exist, and the disappearance of these results might represent a regression to the mean and a consensus view in the financial literature.

Furthermore, we find notable effects regarding the matching of SRI and the respective benchmark. If a study applies one-to-one matching, the performance of SRI relative to the market portfolio decreases by 44.6 percentage points, which is significant at the 5% level. The probability that a study finds SRI outperformance decreases by 25.4% when one-to-one matching is applied, and by 18.1%. when one-to-many matching is applied. However, the coefficients are only weakly statistically significant. For the OLS regression, the coefficients of *one-to-many matching* also has the expected sign. Thus, if matching techniques are applied, SRI, if anything, performs worse than conventional assets. However, matching can have different quality levels and use a different set of criteria, as captured by the variable *NumMatchCriteria*. We find that if more matching criteria are applied (i.e., the more similar SRI and conventional assets are), the better the performance of SRI. If a study is published in a peer-review journal, this increases the performance of SRI relative to the market by 33.1 percentage points. Likewise, publications in a peer-review journal increase the probability that

[21] Of the studies included in the meta-regression, 39 studies and 163 observations are published in a finance journal. Most commonly used performance measures are factor models. In total, 70 observations (24 studies) use multi-factor model alpha and 31 observations (16 studies) one-factor model Jensen's alpha. The second-largest group is risk-adjusted measures (9 studies and 34 observations). Furthermore, 14 observations (7 studies) report an SRI dummy variable. Excess return is used in 6 studies and 11 observations. Last, 2 studies and 3 observations calculate other returns. No study published in a finance journal uses efficiency measures as a performance measure.

a study reports SRI outperformance by 23.7%. Both coefficients of *peer-reviewed* journals are positive and significant at the 1% level. Financial data from *Morningstar* are more likely to report better SRI performance than data from all other databases. If a study reports the ESG taxonomy used, this reduces the performance of SRIs by 31.5 percentage points relative to the market benchmark. The coefficient is significant at the 1% level. Finally, *SIF taxonomy* is more likely to report better SRI performance.

[Tables 4-8 about here]

4.6 Event Studies and Special Types of Studies

We did not consider event studies and other special types of studies in the meta-analyses, because the returns they produce are not comparable with, for example, the performance measures we have considered so far. This is because the sample periods of event studies are typically shorter and focus on particular events. We refer to special types of studies if they consider particular themes, such as the comparison of SRI with sin investing. These studies are rare, which makes it impossible for us to carry out a rigorous statistical meta-analysis. Nevertheless, it is worth mentioning that several event studies show that SRI outperforms conventional investments under particular circumstances. For example, Nakai et al. (2016) examine the effect of the 2008 global financial crisis on Japanese SRI and conventional funds. The event influenced SRI fund performance significantly positively and conventional fund performance significantly negatively when applying a Fama–French three-factor model. The authors also conclude that risk diversification is made difficult because SRI investment possibilities are limited. SRI also weathered the Lehman Brothers bankruptcy better than conventional funds (Nakai et al., 2016). Arefeen and Shimada (2020) also conducted an event study, investigating SRI and conventional funds in the Japanese market. The relevant events chosen for the study were the 2016 US election and Brexit. SRI performed better than conventional funds during the US election when testing performance with EGARCH and Fama–French factor models. The opposite result occurred for Brexit; that is, SRI performed worse than conventional funds.

Paul (2017) and Roofe Sattlethight (2011) split US market data into market expansion and contraction phases between 1991 and 2009 for their event studies. Both authors find overall no difference for SRI and market portfolios. Bhana's (2018) event study examines market response to firm additions and removals from the JSE Socially Responsible Investment Index from 2004 to 2014. Bhana finds that firm additions result in statistically positive abnormal returns. Nitani et al. (2015) report similar results. Another important study is that of Chan and Walter (2014), who examine the effect of investing in environmentally friendly companies and their initial public offerings and seasoned equity offerings. The authors show that these companies outperform in the long run.

So far, we did not consider studies that compare SRI with sin investments or evaluate the performance of sin assets only. Sin assets refer to firms involved in the manufacture of, for example, alcohol, gambling, and tobacco. These assets attract fewer investors and therefore generate higher risk-adjusted returns (Hong & Kacperczyk, 2009). Durand et al. (2013) find that there are cultural differences in holding sin stocks in the Pacific-Basin region. Countries that are culturally more similar to the US, such as Australia, are less likely to favor sin stocks than those that are culturally distant, such as Japan. Steyn and Viviers (2020) show that it is not possible to generate risk-adjusted outperformance with morally questionable investing in South Africa. Weisskopf (2020) reports that investing in cannabis generates extremely high risk and returns. Trinks and Scholtens (2017) find that investing in controversial stocks results in higher risk-adjusted returns. Roofe Sattlethight's (2011) event study compares the performance of SRI with the Vice Fund, which invests in companies that have significant engagement in or derive a substantial portion of their revenues from businesses devoted to behaviors that are traditionally regarded as morally questionable vices. The author finds no significant difference between SRI and the Vice Fund performance. Areal et al. (2013) also compare SRI and morally responsible funds with the Vice Fund and show that in times of low market volatility, the Vice Fund outperforms, while it underperforms in times of high volatility. Soler-Domínguez and Matallín-Sáez (2016) find that in expansion periods, the Vice Fund outperforms the market and SRI funds and vice versa in times of economic crisis.

5 Conclusion

In this article, we perform a meta-analysis to examine the performance of SRI, which after a rigorous literature review includes 153 studies and 1,047 observations of SRI performance. We find that, on average, SRI neither outperforms nor underperforms the respective market benchmarks. One reason for this result could be that while the corporate finance literature shows that certain aspects of CSR, for example, create employee satisfaction (Edmans, 2011), the asset pricing literature is generally less convinced of the ability of ESG investment strategies to outperform market benchmarks (Blanchett, 2010). However, in line with modern portfolio theory, we find that global SRI outperforms regional sub-portfolios. Furthermore, high-quality publications, publications in finance journals, and authors who publish more frequently on SRI are all less likely to report SRI outperformance. In line with the notion that explicitly including more factors in a model removes more unexplained variance, we find that including more factors in a model reduces the likelihood that outperformance of SRI is found.

Our findings contribute not only to the literature on SRI, but also to evaluations of ESG more generally. We provide indirect evidence that academic disciplines, traditions, and the specific methods used can influence how SRI is valued. While within a certain discipline including a simple dummy variable to identify the outperformance of SRI might serve as a valid statistical identification, in other disciplines and sub-disciplines such as theoretical and empirical asset management, hundreds of factors need to be included, to allow for a convincing statistical test of SRI outperformance. Our study deliberately did not follow a narrow path in evaluating only studies that are alike. While our arguable crude measures of SRI performance might be criticized, the virtue of our investigation is that we also provide evidence over a wide range of disciplines, traditions, and specific methods that have been applied to study SRI. Some of these methods are well known to asset management scholars, such as risk-adjusted measures and multi-factor models, while others might not be known to them in the present context, such as DEA outputs.

Finally, our findings also have implications for asset management practice. Not only since the recent introduction of the EU taxonomy has the consideration of ESG factors become indispensable. Many conventional fund managers have integrated ESG factors into their investment strategies for years (Duuren et al., 2016). Aside from moral arguments, there are several economic reasons for companies to consider ESG factors, such as mitigating the impact of crises, the demand of stakeholders to consider environmental or social impacts, or simply aiming to achieve higher profits (Alda, 2020; Revelli, 2017). Furthermore, given that different taxonomies with regard to ESG standards exist (Beerbaum, 2021), investors need to be aware that different taxonomies might lead to different risk-adjusted SRI returns. As such, future empirical research might investigate whether a specific taxonomy has a positive or negative long-term impact on asset performance. From a normative standpoint, it is not clear per se whether investors should be led toward SRI by asset managers or whether they should freely determine their portfolios independent of ESG standards. After all, investors could use additional returns more efficiently to pursue their individual ESG goals.

References

Abdelsalam, O., Duygun, M., Matallín-Sáez, J. C., & Tortosa-Ausina, E. (2014). Do ethics imply persistence? The case of Islamic and socially responsible funds. *Journal of Banking & Finance*, *40*, 182–194.

Aghion, P., & Bolton, P. (1992). An incomplete contracts approach to financial contracting. *Review of Economic Studies*, *59*(3), 473.

Ait El Mekki, O. (2020). Is socially responsible investment outperforming conventional investment or not? A meta-analysis. *American Journal of Industrial and Business Management*, *10*(11), 1760–1784.

Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, *65*(10), 4451–4469.

Alda, M. (2020). ESG fund scores in UK SRI and conventional pension funds: Are the ESG concerns of the SRI niche affecting the conventional mainstream?. *Finance Research Letters*, *36*, 101313.

Alexander, J. C., & Mabry, R. H. (1994). Relative significance of journals, authors, and articles cited in financial research. *Journal of Finance*, *49*(2), 697–712.

Alm, J., & Reed, W. R. (2015). The need for replications. *Public Finance Review*, *43*(2), 139–142.

Amara, N., Landry, R., & Halilem, N. (2015). What can university administrators do to increase the publication and citation scores of their faculty members?. *Scientometrics*, *103*(2), 489–530.

Ang, A., & Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, *106*(1), 132–156.

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

Arefeen, S., & Shimada, K. (2020). Performance and resilience of socially responsible investing (SRI) and conventional funds during different shocks in 2016: Evidence from Japan. *Sustainability*, *12*(2), 540.

Arnold, T., Butler, A. W., Crack, T. F., & Altintig, A. (2003). Impact: What influences finance research? *Journal of Business*, *76*(2), 343–361.

Badía, G., Cortez, M. C, & Ferruz, L. (2020). Socially responsible investing worldwide: Do markets value corporate social responsibility? *Corporate Social Responsibility and Environmental Management*, *27*(6), 2751–2764.

Barber, B. M., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Financial Economics*, *139*(1), 162–185.

Basso, A., & Funari, S. (2001). A data envelopment analysis approach to measure the mutual fund performance. *European Journal of Operational Research*, *135*(3), 477–492.

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

Bauer, R., Otten, R., & Rad, A. T. (2006). Ethical investing in Australia: Is there a financial penalty? *Pacific-Basin Finance Journal*, *14*(1), 33–48.

Bauer, R., Ruof, T., & Smeets, P. (2021). Get real! Individuals prefer more sustainable investments. *Review of Financial Studies, 34*(8), 3976–4043.

Beerbaum, D. O. (2021). Green quadriga? - EU - Taxonomy, non-financial reporting directive, EBA Pillar III ESG risks and IFRS foundation. SSRN Working Paper.

Bhana, N. (2018). Corporate social responsibility initiatives and its impact on firm share price performance: Evidence from South Africa. *Journal of Contemporary Management*, *15*(1), 100–126.

Blanchett, D. M. (2010). Exploring the cost of investing in socially responsible mutual funds: An empirical study. *Journal of Investing*, *19*(3), 93–103.

Bollen, N. P. B. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, *42*(3), 683–708.

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, *142*(2), 517–549.

Borokhovich, K. A., Bricker, R. J., & Simkins, B. J. (2000). An analysis of finance journal impact factors. *Journal of Finance*, *55*(3), 1457–1469.

Borokhovich, K. A., Lee, A. A., & Simkins, B. J. (2011). A framework for journal assessment: The case of the Journal of Banking & Finance. *Journal of Banking & Finance*, *35*(1), 1–6.

Brooks, C., Fenton, E., Schopohl, L., & Walker, J. (2019). Why does research in finance have so little impact? *Critical Perspectives on Accounting*, *58*, 24–52.

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *Review of Financial Studies*, *5*(4), 553–580.

Calma, A. (2017). The “celebrities” in finance: A citation analysis of finance journals. *Studies in Economics and Finance*, *34*(2), 166–182.

Caplan, L., Griswold, J. S., & Jarvis, W. F. (2013). From SRI to ESG: The Changing World of Responsible Investing. *Commonfund Institute*. Retrieved January 3, 2023, from https://files.eric.ed.gov/fulltext/ED559300.pdf.

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

Chan, K. C., Chang, C.-H., & Chang, Y. (2015). The network effects of publishing in finance. *North American Journal of Economics and Finance*, *33*, 305–316.

Chan, P. T., & Walter, T. (2014). Investment performance of “environmentally-friendly” firms and their initial public offers and seasoned equity offers. *Journal of Banking & Finance*, *44*, 177–188.

Chen, N.-F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, *59*(3), 383–403.

Chen, Z., & Knez, P. J. (1996). Portfolio performance measurement: Theory and applications. *Review of Financial Studies*, *9*(2), 511–555.

Chung, H., Lee, H.-H., & Tsai, P.-C. (2012). Are green fund investors really socially responsible? *Review of Pacific Basin Financial Markets and Policies*, *15*(04), 1250023.

Chung, K. H., Cox, R. A. K., & Mitchell, J. B. (2001). Citation patterns in the finance literature. *Financial Management*, *30*(3), 99.

Cornell, B. (2021). ESG preferences, risk and return. *European Financial Management*, *27*(1), 12–19.

Coval, J. D., & Moskowitz, T. J. (1999). Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance*, *54*(6), 2045–2073.

Cumming, D. (2020). Publishing in finance versus entrepreneurship/management journals. In M. Wright, D. J. Ketchen Jr., & T. Clark (Eds.), *How to get published in the best management journals.* Edward Elgar Publishing.

Cumming, D., & Dai, N. (2010). Local bias in venture capital investments. *Journal of Empirical Finance*, *17*(3), 362–380.

Del Miralles-Quirós, M. M., & Miralles-Quirós, J. L. (2017). Improving diversification opportunities for socially responsible investors. *Journal of Business Ethics*, *140*(2), 339–351.

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.

Dorfleitner, G., & Grebler, J. (2022). Corporate social responsibility and systematic risk: international evidence. *Journal of Risk Finance*, *23*(1), 85–120.

Dumrose, M., Rink, S., & Eckert, J. (2022). Disaggregating confusion? The EU Taxonomy and its relation to ESG rating. *Finance Research Letters*, *48*, 102928.

Durand, R. B., Koh, S., & Tan, P. L. (2013). The price of sin in the Pacific-Basin. *Pacific-Basin Finance Journal*, *21*(1), 899–913.

Duuren, E., Plantinga, A., & Scholtens, B. (2016). ESG integration and the investment management process: Fundamental investing reinvented. *Journal of Business Ethics*, *138*(3), 525–533.

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

Eun, C. S., Kolodny, R., & Resnick, B. G. (1991). U.S.-based international mutual funds: A performance evaluation. *Journal of Portfolio Management*, *17*(3), 88–94.

European Fund and Asset Management Association (2021). Asset management in Europe: An overview of the asset management industry. Retrieved January 31, 2022, from https://www.efama.org/newsroom/news/asset-management-report-2021.

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

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*(1), 55–84.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, *128*(2), 234–252.

Ferson, W. E., & Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, *51*(2), 425–461.

French, K., & Poterba, J. M. (1991). Investor diversification and international equity markets. *American Economic Review*, *81*(2), 222–226.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210–233.

Fuyuno, I., & Cyranoski, D. (2006). Cash for papers: Putting a premium on publication. *Nature*, *441*(7095), 792–793.

Gagliardini, P., Ossola, E., & Scaillet, O. (2020). Estimation of large dimensional conditional factor models in finance. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, & R. L. Matzkin (Eds.), *Handbook of econometrics* (Vol. 7A, pp. 219–282). Elsevier.

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

Global Sustainable Investment Alliance (2021). *Global sustainable investment review 2020*. Retrieved January 31, 2022, from http://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf.

Goetzmann, W. N., & Ukhov, A. D. (2006). British investment overseas 1870–1913: A modern portfolio theory approach. *Review of Finance*, *10*(2), 261–300.

Gompers, P., Ishii, J., & Metrick, A. (2003). Corporate governance and equity prices. *Quarterly Journal of Economics*, *118*(1), 107–156.

Graber, M., Wälde, K., & Launov, A. (2008). Publish or perish? The increasing importance of publications for prospective economics professors in Austria, Germany and Switzerland. *German Economic Review*, *9*(4), 457–472.

Grauer, R. R., & Hakansson, N. H. (1987). Gains from international diversification: 1968-85 returns on portfolios of stocks and bonds. *Journal of Finance*, *42*(3), 721–739.

Gregory, A., Matatko, J., & Luther, R. (1997). Ethical unit trust financial performance: Small company effects and fund size effects. *Journal of Business Finance & Accounting*, *24*(5), 705–725.

Grossman, S. J., & Hart, O. D. (1992). An analysis of the principal-agent problem. In G. Dionne & S. E. Harrington (Eds.), *Foundations of insurance economics* (pp. 302–340). Springer.

Grubel, H. G. (1968). Internationally diversified portfolios: Welfare gains and capital flows. *American Economic Review*, *58*(5), 1299–1314.

Guo, H., Wang, B., Qiao, X., & Liu, R. (2016). A review of studies on citations and journal ranking in finance. *Managerial Finance*, *42*(4), 303–311.

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

Harbord, R. M., Egger, M., & Sterne, J. A. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, *25*(20), 3443-3457.

Hardin, W. G., Liano, K., Chan, K. C., & Fok, R. C. W. (2008). Finance editorial board membership and research productivity. *Review of Quantitative Finance and Accounting*, *31*(3), 225–240.

Harris, R. D. F., & Kucukozmen, C. C. (2001). The empirical distribution of UK and US stock returns. *Journal of Business Finance & Accounting*, *28*(5/6), 715–740.
Heinkel, R., Kraus, A., & Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, *36*(4), 431–449.
Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*(4), 513–533.
Hilmer, M. J., & Hilmer, C. E. (2011). Do editors favor their students' work? A test of undue favoritism in top economics journals. *Economics Bulletin*, *31*(1), 53–65.
Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15–36.
Hornuf, L., Schmitt, M., & Stenzhorn, E. (2022). The local bias in equity crowdfunding: Behavioral anomaly or rational preference?. *Journal of Economics & Management Strategy*.
Horváthová, E. (2010). Does environmental performance affect financial performance? A meta-analysis. *Ecological Economics*, *70*(1), 52–59.
Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *Review of Financial Studies*, *28*(3), 650–705.
Hudson, R. (2005). Ethical investing: Ethical investors and managers. *Business Ethics Quarterly*, *15*(4), 641–657.
Humphrey, J. E., & Lee, D. D. (2011). Australian socially responsible funds: Performance, risk and screening intensity. *Journal of Business Ethics*, *102*(4), 519–535.
Hwang, J. H., Kim, D. H., & Chun, S. H. (2011). Corporate social responsibility and its performances: Application to SRI (socially responsible investing) mutual funds. In J. J. Park, L. T. Yang, & C. Lee (Eds.), *Future information technology* (pp. 200-203). Springer.
Ivković, Z., & Weisbenner, S. (2005). Local does as local Is: Information content of the geography of individual investors' common stock investments. *Journal of Finance*, *60*(1), 267–306.
Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, *51*(1), 3–53.
Jarrow, R., & Protter, P. (2016). Positive alphas and a generalized multiple-factor asset pricing model. *Mathematics and Financial Economics*, *10*(1), 29-48.
Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *Journal of Finance*, *23*(2), 389–416.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, *3*(4), 305–360.
Jufang, S., & Huiyun, S. (2011). The outflow of academic papers from China: Why is it happening and can it be stemmed? *Learned Publishing*, *24*(2), 95–97.
Kadiyala, P. (2009). Socially responsible investing: Moral and optimal? *Investment Management and Financial Innovations*, *6*(3), 165–178.

Kiymaz, H. (2012). SRI mutual fund and index performance. In H. K. Baker & J. R. Nofsinger (Eds.), *Socially responsible finance and investing: Financial institutions, corporations, investors, and activists* (pp. 425–442). Wiley.

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

Larcker, D. F., & Watts, E. M. (2020). Where's the greenium? *Journal of Accounting and Economics*, *69*(2/3), 101312.

Lessard, D. R. (1976). World, country, and industry relationships in equity returns: Implications for risk reduction through international diversification. *Financial Analysts Journal*, *32*(1), 32–38.

López-Arceiz, F. J., Bellostas-Pérezgrueso, A. J., & Moneva, J. M. (2018). Evaluation of the cultural environment's impact on the performance of the socially responsible investment funds. *Journal of Business Ethics*, *150*(1), 259–278.

Luo, H. A., & Balvers, R. J. (2017). Social screens and systematic investor boycott risk. *Journal of Financial and Quantitative Analysis*, *52*(1), 365–399.

Maberly, E. D., & Pierce, R. M. (2007). Citation patterns within the leading top-tier finance journals: implications for journal rankings and other issues. SSRN Working Paper.

Macdonald, S., & Kam, J. (2007). Ring a ring o'roses: Quality journals and gamesmanship in management studies. *Journal of Management Studies*, *44*(4), 640–655.

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

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

Mossin, J. (1966). Equilibrium in a capital asset market. Econometrica, 34(4), 768–783.

Mueller, S.A: (1991). The opportunity cost of discipleship: Ethical mutual funds and their performance. *Sociological* Analysis, 52, 111–124.

Nainggolan, Y., How, J., & Verhoeven, P. (2016). Ethical screening and financial performance: The case of Islamic equity funds. *Journal of Business Ethics, 137*(1), 83–99.

Nakai, M., Yamaguchi, K., & Takeuchi, K. (2016). Can SRI funds better resist global financial crisis? Evidence from Japan. *International Review of Financial Analysis, 48*, 12–20.

Nitani, M., Carriere, B., & Bleackley, A. (2015). Recognizing corporate citizenship: Market reactions. *Journal of Sustainable Finance & Investment*, *5*(1/2), 85–102.

Oltheten, E., Theoharakis, V., & Travlos, N. G. (2003). Faculty perceptions and readership patterns of finance journals: A global view. *SSRN Working Paper.*

Pasewark, W. R., & Riley, M. E. (2010). It's a matter of principle: The role of personal values in investment decisions. *Journal of Business Ethics*, *93*(2), 237–253.

Pastor, L., Stambaugh, R., & Taylor, L. (2021a). Dissecting green returns. *National Bureau of Economic Research.*

Pastor, L., Stambaugh, R., & Taylor, L. (2021b). Sustainable investing in equilibrium. *Journal of Financial Economics 142*(2), 550-571.

Paul, K. (2017). The effect of business cycle, market return and momentum on financial performance of socially responsible investing mutual funds. *Social Responsibility Journal*, *13*(3), 513–528.

Pedersen, L. H., Fitzgibbons, S., & Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, *142*(2), 572–597.

Powdthavee, N., Riyanto, Y. E., & Knetsch, J. L. (2018). Lower-rated publications do lower academics' judgments of publication lists: Evidence from a survey experiment of economists. *Journal of Economic Psychology*, *66*, 33–44.

Rathner, S. (2013). The influence of primary study characteristics on the performance differential between socially responsible and conventional investment funds: A meta-analysis. *Journal of Business Ethics*, *118*(2), 349–363.

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

Renneboog, L., Ter Horst, J., & Zhang, C. (2011). Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation*, *20*(4), 562–588.

Revelli, C. (2017). Socially responsible investing (SRI): From mainstream to margin?. *Research in International Business and Finance*, *39*, 711–717.

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

Roofe Sattlethight, A. (2011). The effect of the business cycle on the performance of socially responsible equity mutual funds. Working paper, Florida International University.

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

Rubin, A., & Rubin, E. (2021). Systematic bias in the progress of research. *Journal of Political Economy*, *129*(9), S. 2666–2719.

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

Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. *Science Advances*, *7*(21), eabd1705.

Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, *59*(5), 1045–1061.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, *39*(1), 119–138.

Solberg, L. (2022). Why sustainable strategies outperformed in 2021. Morningstar, Inc. Retrieved February 10, 2022, from https://www.morningstar.com/articles/1075190/why-sustainable-strategies-outperformed-in-2021.

Soler-Domínguez, A., & Matallín-Sáez, J. C. (2016). Socially (ir)responsible investing? The performance of the VICEX Fund from a business cycle perspective. *Finance Research Letters*, *16*, 190–195.

Solnik, B. H. (1974). Why not diversify internationally rather than domestically? *Financial Analysts Journal*, *30*(4), 48–54.

Steyn, J. P., & Viviers, S. (2020). The performance of morally questionable shares in South Africa (2004–2019). *Journal of Economic and Financial Sciences*, *13*(1), 12.

Stone, M., & Rasp, J. (1991). Tradeoffs in the choice between logit and OLS for accounting choice studies. *Accounting Review*, *66*(1), 170–187.

Swidler, S., & Goldreyer, E. (1998). The value of a finance journal publication. *Journal of Finance*, *53*(1), 351–363.

Treynor, J. (1965). How to rate management of investment funds. *Harvard Business Review*, *43*(1), 63–75.

Treynor, J., & Mazuy, K. (1966). Can mutual funds outguess the market? *Harvard Business Review*, *44*(4), 131–136.

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

Tully, S. M., & Winer, R. S. (2014). The role of the beneficiary in willingness to pay for socially responsible products: A meta-analysis. *Journal of Retailing*, *90*(2), 255–274.

Wallis, M. v., & Klein, C. (2015). Ethical requirement and financial interest: a literature review on socially responsible investing. *Business Research, 8*(1), 61–98.

Weisskopf, J.-P. (2020). Breaking bad: An investment in cannabis. *Finance Research Letters*, *33*, 101201.

Zerbib, O. D. (2020). A sustainable capital asset pricing model (S-CAPM). Evidence from green investing and sin stock exclusion. Working paper, Tilburg University.

Figure 1: Consideration of SRI and ESG in the Sample Studies

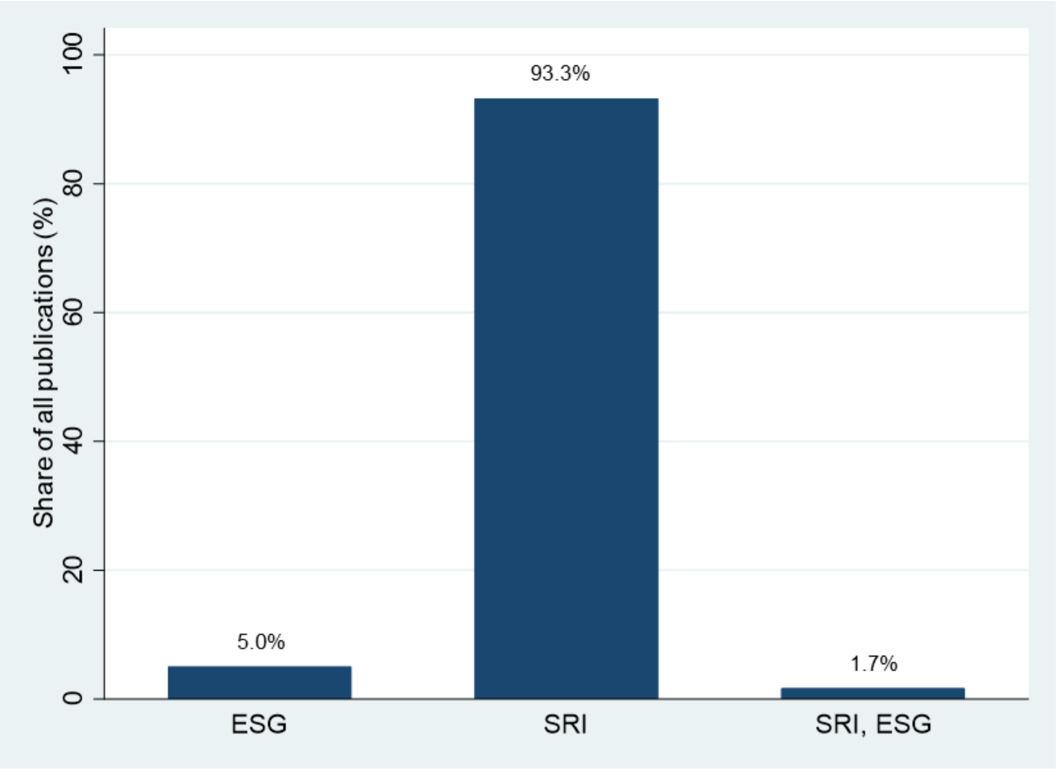

Figure 2: Consideration of SRI and ESG in the Sample Studies Over Time

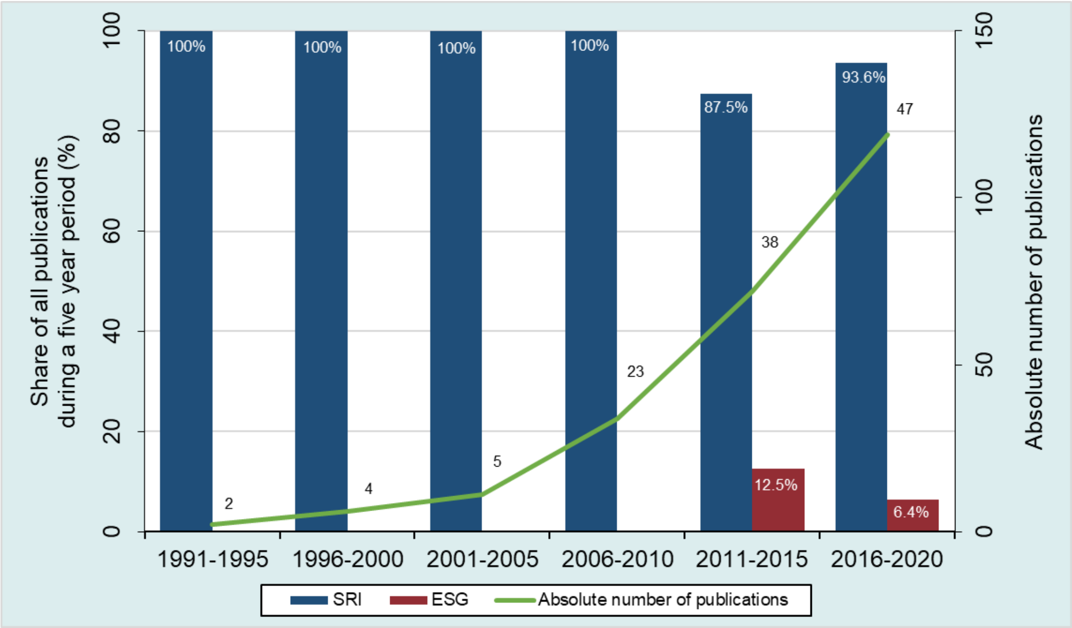

Figure 3: Funnel Plot Reporting the Performance Difference Between SRI and the Market Portfolio in the Respective Study and Sample Size for SRI Funds (n=413)

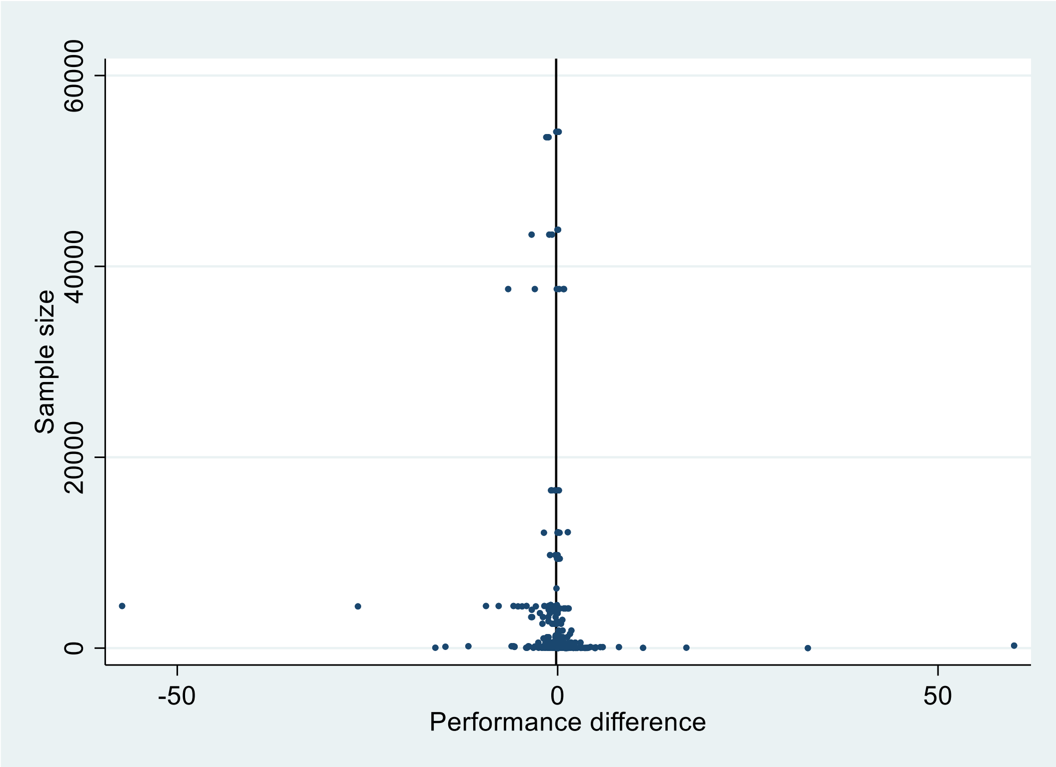

Figure 4: Funnel Plot Reporting the Performance Difference Between SRI and the Market Portfolio in the Respective Study and Sample Size for SRI Stocks (n=28)

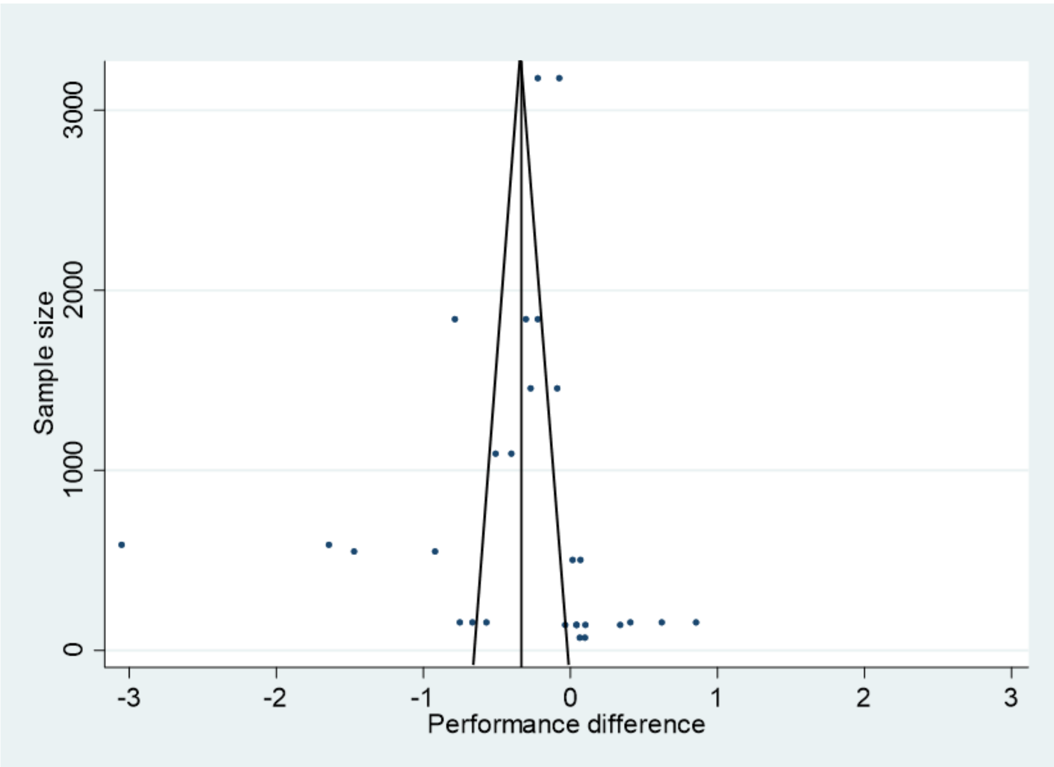

Table 1: Performance Evaluation SRI Without Market Comparison

Performance measure	Studies	Number of observations (n)	Mean	Std. Dev.	Min	Max
Annual returns funds	21	47	0.038	0.091	-0.245	0.184
Annual returns stocks	2	8	0.232	0.100	0.113	0.360
Annual returns bonds	2	23	0.070	0.008	0.051	0.082
Excess returns funds	2	4	-0.004	0.014	-0.018	0.016
Excess returns stocks	4	18	0.012	0.030	-0.012	0.070
Annual i-spread bonds	1	1	47.539	.	47.539	47.539
Risk-adjusted measures funds	9	22	0.078	0.946	-1.984	2.508
Risk-adjusted measures stocks	3	19	0.824	0.349	0.240	1.490
Other returns funds	2	2	0.007	0	0.007	0.007
Factor models funds	30	139	-0.322	2.981	-14.112	18.900
Factor models stocks	12	59	-0.014	0.223	-1.138	0.408
Factor models bonds	1	8	0.006	0.008	-0.003	0.017
Efficiency measures funds	3	3	1.067	0.352	0.666	1.324

Notes: This table provides descriptive statistics for the performance measures of SRI funds, stocks, and bonds. The first column shows the number of studies and the second column the number of observations in our dataset that do not conduct a market comparison. We included multiple observations for a study when the study used multiple independent regions, independent sampling periods, or more than one method of performance evaluation.

Table 2: Performance Evaluation SRI versus Market Portfolio

Performance measure	N	n	Mean	Std. Dev.	Min	Max	z-score	p-value
Funds								
Panel A: Annual returns								
(1) SRI funds	65	114	0.061	0.075	-0.435	0.326		
(2) Market portfolio	65	114	0.063	0.066	-0.336	0.270		
(1) – (2)	66	115	-0.002	0.036	-0.138	0.161	0.524	0.600
(3) Performance difference	65	114	0.515	5.685	-4.241	60.000		0.057
(4) Outperformance of SRI	66	115	0.391	0.490	0	1	2.331	0.020
Panel B: Excess returns								
(1) SRI funds	14	31	0.012	0.037	-0.092	0.081		
(2) Market portfolio	14	31	0.027	0.043	-0.094	0.086		
(1) – (2)	15	33	-0.018	0.037	-0.082	0.072	2.078	0.038
(3) Performance difference	14	31	-0.278	1.543	-5.663	4.057		0.011
(4) Outperformance of SRI	15	33	0.242	0.435	0	1	-2.959	0.003
Panel C: Risk-adjusted measures								
(1) SRI funds	27	81	-0.021	6.372	-50.283	25.250		
(2) Market portfolio	27	81	0.005	5.510	-43.933	20.390		
(1) – (2)	28	83	-0.026	1.242	-6.350	4.860	0.152	0.879
(3) Performance difference	27	81	0.614	2.531	-3.400	16.920		0.738
(4) Outperformance of SRI	28	83	0.506	0.503	0	1	0.110	0.913
Panel D: Other returns								
(1) SRI funds	4	5	-0.042	0.298	-0.551	0.220		
(2) Market portfolio	4	5	-0.016	0.202	-0.360	0.146		
(1) – (2)	4	5	-0.026	0.004	-0.005	0.005	1.497	0.134
(3) Performance difference	4	5	-0.130	0.408	-0.530	0.500		0.375
(4) Outperformance of SRI	5	7	0.143	0.378	0	1	-1.890	0.059
Panel E: Factor models								
(1) SRI funds	57	232	-0.019	0.393	-1.667	2.520		
(2) Market portfolio	57	232	-0.038	0.368	-2.036	2.016		
(1) – (2)	59	236	0.018	0.234	-1.136	1.780	2.065	0.039
(3) Performance difference	56	229	-3.306	29.254	-397.000	32.893		0.077
(4) Outperformance of SRI	59	236	0.411	0.493	0	1	-2.734	0.006
Panel F: Efficiency measures								
(1) SRI funds	6	15	0.492	0.254	0.117	0.952		
(2) Market portfolio	6	15	0.513	0.215	0.150	0.972		
(1) – (2)	6	15	-0.021	0.109	-0.270	0.141	-0.021	0.983
(3) Performance difference	6	15	-0.074	0.212	-0.519	0.232		0.302
(4) Outperformance of SRI	6	15	0.333	0.488	0	1	-1.291	0.197
Panel G: Dummy variables								
(1) SRI funds	12	22	-1.296	9.705	-26.400	14.400		
(2) Market portfolio	12	22	14.469	22.430	-8.712	61.2		
(1) – (2)	12	22	-15.765	26.0512	-87.6	11.652	2.324	0.020
(3) Performance difference	12	22	-1.115	2.905	-9.219	4.928		0.078
(4) Outperformance of SRI	12	22	0.273	0.456	0	1	-2.132	0.033
Stocks								
Panel H: Annual returns								
(1) SRI stocks	19	50	0.118	0.090	-0.027	0.334		
(2) Market portfolio	19	50	0.103	0.081	0.011	0.327		
(1) – (2)	19	50	0.015	0.063	-0.220	0.161	-0.969	0.333
(3) Performance difference	19	50	0.293	0.957	-3.052	3.628		0.044
(4) Outperformance of SRI	19	50	0.640	0.485	0	1	1.980	0.048
Panel I: Excess returns								

(1) SRI stocks	4	11	0.166	0.092	0.020	0.341		
(2) Market portfolio	4	11	0.092	0.062	-0.014	0.198		
(1) – (2)	4	11	0.074	0.061	0.008	0.204	-2.003	0.045
(3) Performance difference	4	11	1.149	1.062	0.041	3.286		0.001
(4) Outperformance of SRI	4	11	1.000	0.000	1	1	3.317	0.001
Panel J: Risk-adjusted measures								
(1) SRI stocks	16	59	0.923	1.202	0.006	5.127		
(2) Market portfolio	16	59	0.593	0.730	-0.416	3.371		
(1) – (2)	16	59	0.330	0.804	-1.136	3.523	-1.922	0.055
(3) Performance difference	16	59	31.676	174.416	-0.786	1195.455		0.000
(4) Outperformance of SRI	16	59	0.729	0.448	0	1	3.515	0.000
Panel K: Factor models								
(1) SRI stocks	9	31	0.155	0.276	-0.028	1.092		
(2) Market portfolio	9	31	0.185	0.358	-0.006	1.663		
(1) – (2)	9	31	-0.030	0.254	-1.109	0.310	0.254	0.799
(3) Performance difference	9	27	-8.006	24.994	-86.500	19.568		0.648
(4) Outperformance of SRI	9	31	0.290	0.461	0	1	-2.335	0.020
Panel L: Dummy variables								
(1) SRI stocks	1	4	28.791	27.742	-11.592	50.652		
(2) Market portfolio	1	4	21.042	39.628	-28.980	67.788		
(1) – (2)	1	4	7.749	16.629	-17.136	17.388	-0.577	0.564
(3) Performance difference	1	4	0.430	0.459	-0.253	0.744		0.625
(4) Outperformance of SRI	1	4	0.750	0.500	0	1	1.000	0.317
Bonds								
Panel M: Annual returns								
(1) SRI bonds	3	5	0.057	0.024	0.030	0.091		
(2) Market portfolio	3	5	0.050	0.023	0.025	0.082		
(1) – (2)	3	5	0.008	0.004	0.004	0.014	-0.733	0.463
(3) Performance difference	3	5	0.180	0.125	0.059	0.378		0.063
(4) Outperformance of SRI	3	5	1.000	0.000	1	1	2.236	0.025
Panel N: Mean spread								
(1) – (2)	1	1	0.280	.	0.280	0.280		
Panel O: Risk-adjusted measures								
(1) SRI bonds	1	2	0.490	0.057	0.450	0.530		
(2) Market portfolio	1	2	0.405	0.106	0.330	0.480		
(1) – (2)	1	2	0.085	0.049	0.050	0.120	-0.775	0.439
(3) Performance difference	1	2	0.234	0.183	0.104	0.364		0.500
(4) Outperformance of SRI	1	2	1.000	0.000	1	1	1.414	0.157
Panel P: Factor models								
(1) SRI bonds	3	15	0.001	0.010	-0.011	0.016		
(2) Market portfolio	3	15	-0.005	0.011	-0.021	0.015		
(1) – (2)	3	15	0.006	0.006	-0.001	0.014	-1.847	0.065
(3) Performance difference	3	15	0.528	0.535	-0.078	1.259		0.001
(4) Outperformance of SRI	3	15	0.933	0.258	0	1	3.357	0.001
Panel Q: Dummy variables								
(1) SRI bonds	3	5	0.467	0.536	0.020	1.340		
(2) Market portfolio	3	5	0.276	0.383	-0.160	0.800		
(1) – (2)	3	5	0.191	0.346	-0.230	0.560	-0.731	0.465
(3) Performance difference	3	4	0.834	1.895	-0.920	3.500		0.625
(4) Outperformance of SRI	3	5	0.800	0.447	0	1	1.342	0.180

Notes: This table provides descriptive statistics for different performance measures of SRI funds, stocks, and bonds and the market portfolio. *N* is the number of studies and *n* the number of observations in the dataset. We included multiple observations for a study when the study used multiple independent regions, independent sampling periods, or more than one method of performance evaluation. Performance measures are annualized. Annual returns are arithmetic raw returns. Other returns (Panel D) include four-factor adjusted abnormal returns, excess standard deviation–adjusted returns, and DEA outputs. Efficiency measures (Panel F) combine efficiency scores of DEA and FDH. For each performance measure, we report the performance difference between SRI and the market portfolio (1) – (2). To test whether the difference between SRI and the market is statistically significant, we conduct a Wilcoxon rank-sum test. Moreover, in line (3), we report the average percentage differences in performance of SRI relative to the market portfolio. Note that the number of observations might differ from (1) – (2), because a few studies only report the performance difference but not the performance of SRI and the market portfolio. To test whether the *performance difference* is statistically different from zero, we conduct a sign test. *Outperformance of SRI* is the share of studies reporting SRI outperformance. To test whether significantly more than 50% of the observations report SRI outperformance, we employ a Wilcoxon signed-rank test.

Table 3: Summary Statistics

Dependent Variables	n	Mean	Median	Std. Dev.	Min	Max
Performance difference	648	0.015	0	0.562	-3.970	11.955
Outperformance of SRI	666	0.453	0	0.498	0	1
Continuous Explanatory Variables						
Impact factor	666	1.629	0	2.295	0	12.110
Number of publications	666	2.350	2	1.756	1	7
Number of matching criteria	666	0.574	0	1.226	0	5
Dichotomous Explanatory Variables		Mean	Variable = 1			
Region						
a. Asia		0.173	115			
b. Europe		0.366	244			
c. Global		0.027	18			
d. Rest of the world		0.147	98			
e. US		0.371	247			
Journal information						
a. Finance journal		0.311	207			
b. Peer-reviewed journal		0.764	509			
Performance evaluation						
a. Conditional evaluation		0.077	51			
b. Mean returns		0.248	165			
c. Excess returns		0.066	44			
d. Risk-adjusted measures		0.213	142			
e. Other returns		0.011	7			
f. One-factor model Jensen’s alpha		0.120	80			
g. Multi-factor model alpha		0.281	187			
h. Efficiency measures		0.023	15			
i. SRI dummy variables		0.039	26			
Sample period						
a. 1981–1990		0.029	19			
b. 1991–2000		0.303	202			
c. 2001–2010		0.472	314			
d. 2011–2020		0.216	144			
Matching procedure						
a. One-to-one		0.065	43			
b. One-to-many		0.173	115			
c. No matching		0.762	508			
Investment type						
a. Fund		0.767	511			
b. Stock		0.233	155			
Benchmark						
a. Same asset class (funds or stocks)		0.799	532			
b. Index		0.201	134			
Survivorship bias considered		0.482	321			
Financial databases						
a. Bloomberg		0.105	70			
b. CRSP		0.189	126			
c. Morningstar		0.177	118			
d. Thomson Reuters		0.398	265			
e. Other		0.338	225			
ESG taxonomy		0.605	403			
ESG databases						
a. Bloomberg taxonomy		0.039	26			
b. Morningstar taxonomy		0.155	103			
c. SIF taxonomy		0.122	81			

d.	SocialFunds.com taxonomy	0.029	19
e.	Vigeo taxonomy	0.069	46
f.	Other taxonomy	0.297	198
Number of observations			666

Table 4: Meta-Regression Funds and Stocks

	(1) OLS	(2) Probit		(1) OLS	(2) Probit
Asia	0.019	0.073	NumMatchCriteria	0.159***	0.071**
	(0.104)	(0.064)		(0.053)	(0.032)
Europe	-0.173	-0.055	Peer-reviewed	0.331***	0.237***
	(0.109)	(0.065)		(0.101)	(0.060)
Rest of the world	-0.165	-0.101	Stock	0.364***	0.169***
	(0.106)	(0.064)		(0.094)	(0.053)
US	-0.225**	-0.113*	Survivorship bias	0.052	0.070
	(0.098)	(0.060)		(0.086)	(0.050)
Impact factor	-0.094***	-0.045***			
	(0.018)	(0.011)	*Financial databases*		
Finance journal	-0.348***	-0.226***	Bloomberg	-0.118	-0.077
	(0.095)	(0.055)		(0.124)	(0.066)
Number of	-0.047**	-0.034***	CRSP	-0.039	0.114
publications	(0.021)	(0.012)		(0.139)	(0.076)
Excess returns	0.057	0.035	Morningstar	0.398***	0.165**
	(0.117)	(0.080)		(0.127)	(0.071)
Risk-adjusted	0.163**	0.117**	Thomson Reuters	0.049	0.006
measures	(0.077)	(0.054)		(0.088)	(0.050)
Other returns	0.137	-0.248			
	(0.283)	(0.245)	ESG taxonomy	-0.315***	0.028
One-factor model	-0.086	0.073		(0.091)	(0.052)
Jensen's alpha	(0.126)	(0.064)			
Multi-factor model	-0.222**	-0.009	*ESG databases*		
alpha	(0.092)	(0.055)	Bloomberg	-0.007	0.029
Efficiency measures	0.004	-0.093	taxonomy	(0.203)	(0.113)
	(0.161)	(0.131)	Morningstar	0.130	0.062
SRI dummy	-0.247	-0.062	taxonomy	(0.109)	(0.066)
	(0.201)	(0.102)	SIF taxonomy	0.325**	0.082
SPeriod 1981–2000	0.425***	0.131*		(0.134)	(0.079)
	(0.111)	(0.069)	SocialFunds.com	-0.024	-0.113
SPeriod 2001–2010	0.081	-0.051	taxonomy	(0.228)	(0.154)
	(0.087)	(0.058)	Vigeo taxonomy	0.049	0.144*
Conditional	0.189	0.060		(0.154)	(0.085)
evaluation	(0.156)	(0.077)			
Index benchmark	0.269***	0.181***	Constant	-0.012	
	(0.089)	(0.056)		(0.150)	
One-to-one	-0.446**	-0.254*	Observations	648	666
matching	(0.226)	(0.141)	R^2	0.240	
One-to-many	-0.191	-0.181*	Pseudo-R^2		0.125
matching	(0.159)	(0.094)	Mean VIF	2.06	2.07

Notes: This table considers funds and stocks. It shows the results of the OLS (column 1) and the probit models (column 2). The dependent variable of the OLS model is *performance difference*, the percentage performance difference of SRI relative to the market portfolio. The dependent variable of the probit model is *outperformance of SRI*, which takes the value one if the difference between SRI minus market is positive and zero otherwise. Note that some studies only report the differences between the performances of SRI and the market portfolio, which can be used as dependent variable for the probit model. However, we cannot calculate the percentage performance difference from this data, which is why the number of observations differs for column (1) and column (2). The table reports robust standard errors in parentheses and average marginal effects for the probit model. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 5: Meta-Regression Funds

	(1) OLS	(2) Probit		(1) OLS	(2) Probit
Asia	-0.065	0.012	One-to-many	-0.251	-0.207**
	(0.149)	(0.084)	matching	(0.168)	(0.094)
Europe	-0.141	-0.057	NumMatchCriteria	0.155***	0.063**
	(0.155)	(0.084)		(0.054)	(0.032)
Rest of the world	-0.266*	-0.200**	Peer-reviewed	0.202	0.185**
	(0.148)	(0.085)		(0.148)	(0.080)
US	-0.132	-0.079	Survivorship bias	-0.030	0.022
	(0.138)	(0.079)		(0.104)	(0.059)
Impact factor	-0.087***	-0.035***			
	(0.021)	(0.013)	*Financial databases*		
Finance journal	-0.234**	-0.201***	Bloomberg	-0.047	-0.035
	(0.116)	(0.066)		(0.135)	(0.081)
Number of	-0.068***	-0.048***	CRSP	-0.089	0.069
publications	(0.023)	(0.013)		(0.166)	(0.089)
Excess returns	-0.051	-0.101	Morningstar	0.365**	0.113
	(0.147)	(0.101)		(0.149)	(0.079)
Risk-adjusted	0.012	0.104	Thomson Reuters	0.047	-0.042
measures	(0.103)	(0.068)		(0.118)	(0.062)
Other returns	0.023	-0.281			
	(0.294)	(0.240)	ESG taxonomy	-0.256**	0.074
One-factor model	-0.171	0.060		(0.121)	(0.063)
Jensen's alpha	(0.132)	(0.068)			
Multi-factor model	-0.277***	0.000	*ESG databases*		
alpha	(0.100)	(0.063)	Bloomberg	-0.478	-0.274
Efficiency measures	0.002	-0.149	taxonomy	(0.311)	(0.168)
	(0.168)	(0.131)	Morningstar	0.099	0.038
SRI dummy	-0.464*	-0.140	taxonomy	(0.131)	(0.073)
	(0.237)	(0.118)	SIF taxonomy	0.302**	0.041
SPeriod 1981–2000	0.412***	0.167**		(0.142)	(0.083)
	(0.126)	(0.078)	SocialFunds.com	-0.077	-0.128
SPeriod 2001–2010	0.144	0.025	taxonomy	(0.222)	(0.146)
	(0.121)	(0.077)	Vigeo taxonomy	0.033	0.155*
Conditional	0.166	0.057		(0.165)	(0.091)
evaluation	(0.165)	(0.080)			
Index benchmark	0.009	0.000	Constant	0.184	
	(0.122)	(0.074)		(0.202)	
One-to-one	-0.549**	-0.304**	Observations	497	511
matching	(0.234)	(0.141)	R^2	0.196	
			Pseudo-R^2		0.104
			Mean VIF	2.25	2.26

Notes: This table considers funds only. It shows the results of the OLS (column 1) and the probit models (column 2). The dependent variable of the OLS model is *performance difference*, the percentage performance difference of SRI relative to the market portfolio. The dependent variable of the probit model is *outperformance of SRI*, which takes the value one if the difference between SRI minus market is positive and zero otherwise. Note that some studies only report the differences between the performances of SRI and the market portfolio, which can be used as dependent variable for the probit model. However, we cannot calculate the percentage performance difference from this data, which is why the number of observations differs for column (1) and column (2). The table reports robust standard errors in parentheses and average marginal effects for the probit model. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 6: Meta-Regression Stocks

	(1) OLS	(2) Probit
Asia	0.157	0.065
	(0.173)	(0.098)
Europe	-0.513***	-0.325***
	(0.180)	(0.107)
Rest of the world	-0.399**	-0.187*
	(0.186)	(0.098)
US	-0.274*	-0.172*
	(0.152)	(0.089)
Impact factor	-0.097**	0.018
	(0.046)	(0.023)
Finance journal	-0.653***	-0.070
	(0.206)	(0.107)
Number of publications	0.391**	0.331***
	(0.193)	(0.090)
Risk-adjusted measures	0.308**	0.019
	(0.120)	(0.067)
One-factor model Jensen's alpha	0.215	-0.014
	(0.375)	(0.148)
Multi-factor model alpha	0.036	-0.086
	(0.165)	(0.091)
SRI dummy	-0.370	-0.396*
	(0.328)	(0.225)
SPeriod 1981–2000	-0.491	-0.258
	(0.399)	(0.226)
SPeriod 2001–2010	0.430**	0.148
	(0.204)	(0.100)
Peer-reviewed	0.454**	0.135
	(0.221)	(0.103)
Index benchmark	0.580***	0.341***
	(0.152)	(0.080)
Survivorship bias	0.059	0.053
	(0.217)	(0.109)
Constant	-0.633*	
	(0.364)	
Observations	151	155
R^2	0.405	
Pseudo-R^2		0.331
Mean VIF	2.22	2.22

Notes: This table considers stocks only. It shows the results of the OLS (column 1) and the probit models (column 2). The dependent variable of the OLS model is *performance difference*, the percentage performance difference of SRI relative to the market portfolio. The dependent variable of the probit model is *outperformance of SRI*, which takes the value one if the difference between SRI minus market is positive and zero otherwise. Note that some studies only report the differences between the performances of SRI and the market portfolio, which can be used as dependent variable for the probit model. However, we cannot calculate the percentage performance difference from this data, which is why the number of observations differs for column (1) and column (2). The table reports robust standard errors in parentheses and average marginal effects for the probit model. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 7: Meta-Regression Bonds

	(1) OLS
US	-0.570**
	(0.232)
Impact factor	-0.429**
	(0.196)
Number of publications	-0.077
	(0.055)
One-factor model Jensen's alpha	-0.012
	(0.223)
Multi-factor model alpha	0.405*
	(0.197)
SPeriod 2011–2020	-0.602**
	(0.219)
Constant	1.760***
	(0.586)
Observations	26
R^2	0.328
Mean VIF	2.26

Notes: This table shows the results of the OLS model. The dependent variable of the OLS model is *performance difference*, the percentage performance difference of SRI relative to the market portfolio. The table reports robust standard errors in parentheses. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Table 8: Regression on Impact Factor and Finance Journal

Dependent variable	(1) OLS Impact factor	(2) Probit Finance journal
Excess returns	1.231***	0.092
	(0.453)	(0.104)
Risk-adjusted measures	-0.543*	0.080
	(0.322)	(0.085)
Other returns	-0.029	0.263
	(0.869)	(0.181)
One-factor model Jensen's alpha	-0.018	0.238***
	(0.313)	(0.089)
Multi-factor model alpha	0.569*	0.230***
	(0.322)	(0.081)
SRI dummy variables	0.145	0.359***
	(0.463)	(0.111)
Constant	1.512***	
	(0.259)	
Observations	501	501
R^2	0.060	
Pseudo-R^2		0.035
Mean VIF	2.10	2.10

Notes: This table shows the results of the OLS (column 1) and the probit models (column 2). The dependent variable *impact factor* is the journal impact factor; *finance journal* takes the value one if a study is published in a finance journal and zero otherwise. The table reports robust standard errors in parentheses and average marginal effects of the independent variables of the probit model. ***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Appendix. Detailed Explanation of Performance Measures

A commonly used performance measure is mean excess return. The excess return is the return earned by an asset over the risk-free rate of return. A risk-free rate of return considers little to no market risk, like government bonds.

$$ER = R_i - R_f, \quad (1)$$

where R_i are the periodic asset returns, and R_f the risk-free rate.

Mean or excess returns do not consider risk. Therefore, the traditional finance view according to Markowitz (1952) suggests that rational investors can construct optimal portfolios that maximize the expected return given market risk. By comparison, conventional theory states that diversification enables the maximization of returns. The next measure for asset performance is the Sharpe ratio (SR), or the reward-to-variability ratio. It measures the expected excess return of an asset over its risk described by the standard deviation (Sharpe, 1966).

$$SR = \frac{R_i - R_f}{\sigma_i}, \quad (2)$$

where R_i is the mean return of an asset, R_f is the mean return of a risk-free asset, and σ_i is the standard deviation of portfolio returns.

The Treynor ratio (TR) considers the systematic risk between an asset and the market instead of the standard deviation (Treynor, 1965).

$$TR = \frac{R_i - R_f}{\beta_i}, \quad (3)$$

where β_i is the systematic risk.

The information ratio (IR) considers manager ability and therefore divides Jensen's alpha (abnormal return) by the unsystematic risk of the asset.

$$IR = \frac{\alpha_i}{\sigma(e_i)}, \quad (4)$$

where α_i is the Jensen's α of the asset and $\sigma(e_i)$ is the unsystematic risk.

The Modigliani Modigliani (M^2) ratio is a transformation of the SR by considering the difference between asset risk-adjusted returns and that of the market, suggesting that volatilities are the same.

$$M^2 = \frac{R_i - R_f}{\sigma_i} * \sigma_m, \quad (5)$$

where σ_m is the standard deviation of market index.

The Sortino ratio (SoR) is a variation of the SR. It considers downside deviation, which is the standard deviation of negative returns.

$$SoR = \frac{R_i - R_f}{\sigma_d}, \quad (6)$$

where σ_d is the downside risk of an asset.

The last risk-adjusted measure is the omega ratio (OR). The OR weights gains and losses given a minimum return level, the so-called minimum acceptable return.

$$OR = \frac{\int_r^{\infty} \left(1 - F(x)\right) dx}{\int_{-\infty}^{r} F(x) dx}, \quad (7)$$

where F is the cumulative distribution function of returns and r the minimum acceptable return, which is a gain or loss.

We combine these listed ratios into risk-adjusted measures.

Although Sharpe's and Treynor's ratio and Jensen's alpha are based on the CAPM, we treat Jensen's alpha as part of the category factor to compare alphas of factor models. The CAPM, which was developed by Sharpe (1964), Lintner (1965), and Mossin (1966), is an extension of the simple mean-variance model of portfolio selection. The expected asset returns depend on a single factor, which is the beta-coefficient that is the systematic risk factor.

$$E(R_i) = R_f + \beta_i \left(E(R_M) - R_f\right), \quad (8)$$

where $E(R_i)$ is the expected return of an asset, R_f is the risk-free rate, β_i is the beta of an asset, $E(R_M)$ is the expected return of the market portfolio, and $E(R_M) - R_f$ is the market risk premium.

Jensen (1968) further developed Sharpe's and Treynor's ratios to a risk-adjusted model for stock selection by adding fund manager performance. Expected returns from the CAPM are changed into actual realized returns. Jensen's alpha (intercept) measures the additional return a fund manager generates.

$$R_i - R_f = \alpha_i + \beta_i \left(R_M - R_f\right) + \varepsilon_i, \quad (9)$$

where ε_i is the error term.

To explain the cross-sectional variation in asset returns, a single-factor model is not enough. Therefore, multi-factor models were introduced. These models capture non-market but economic influences. There are several extensions by adding new factors, inter alia, to conditional versions. Most commonly used multi-factor models are the three- and four-factor models.

Fama and French (1993) added the relative market value (size: small-cap stock returns minus large-cap stock returns) of a firm and the ratio of book value to market price (high BV/MV stock returns minus low BV/MV stock returns) to the CAPM for the three-factor model.

$$R_i - R_f = \alpha_i + \beta_1 \left(R_M - R_f\right) + \beta_2 SMB + \beta_3 HML + \varepsilon_i, \quad (10)$$

where SMB is the size premium, HML is the value premium factor, and $\beta_{1,2,3}$ are factor coefficients.

Carhart (1997) extends the three-factor model by introducing a four-factor model that contains the momentum factor (UMD, winners minus losers).

$$R_i - R_f = \alpha_i + \beta_1(R_M - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 UMD + \varepsilon_i. \qquad (11)$$

Additional common models that need to be considered are market-timing models of Treynor and Mazuy (1966) and Henriksson and Merton (1981). These approaches examine fund managers' market timing ability.

$$TM = R_i - R_f = \alpha_i + \beta_i(R_M - R_f) + \gamma_i(R_M - R_f)^2 + \varepsilon_i, \qquad (12)$$

where γ_i states market timing, that is, a positive and significant coefficient indicating a fund manager who is a fruitful market timer.

$$HM = R_i - R_f = \alpha_i + \beta_i(R_M - R_f) + \delta_i(R_M - R_f)D + \varepsilon_i \qquad (13)$$

where δ_i indicates market timing and D is a dummy with a value one if market returns are positive and zero otherwise.

Another category for fund performance is the so-called efficiency measure. DEA uses input variables, such as risk levels or investment costs, and output variables, such as mean returns, to measure fund performance and evaluate relative efficiency of decision-making units (Basso & Funari, 2001). FDH is also part of this category, but there are no observations in case of funds. FDH is a non-convex version of DEA (Abdelsalam et al., 2014).

Table A1: Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) Performance difference	1.000																	
(2) Outperformance of SRI	0.098*	1.000																
(3) Asia	0.067	0.166*	1.000															
(4) Europe	0.007	-0.123*	-0.257*	1.000														
(5) Rest of the World	-0.012	-0.012	0.001	-0.281*	1.000													
(6) US	-0.055	-0.025	-0.260*	-0.487*	-0.187*	1.000												
(7) Impact factor	-0.028	-0.115*	-0.080*	0.265*	0.008	-0.041	1.000											
(8) Finance journal	-0.020	-0.071	-0.144*	-0.060	0.014	0.169*	-0.077*	1.000										
(9) number of publications	-0.001	-0.159*	-0.227*	0.227*	-0.025	-0.080*	0.126*	0.012	1.000									
(10) Excess returns	-0.007	-0.012	-0.042	0.024	-0.042	-0.004	0.129*	-0.035	0.136*	1.000								
(11) Risk-adjusted measures	0.114*	0.152*	0.179*	-0.130*	0.063	-0.020	-0.150*	-0.080*	-0.229*	-0.138*	1.000							
(12) Other returns	-0.003	-0.064	-0.047	-0.048	-0.043	0.104*	-0.007	0.026	-0.054	-0.027	-0.054	1.000						
(13) One-factor model	-0.014	0.044	-0.047	-0.041	0.003	0.051	-0.022	0.061	-0.092*	-0.098*	-0.192*	-0.038	1.000					
(14) Multi-factor model alpha	-0.074	-0.133*	-0.126*	0.149*	-0.033	-0.079*	0.075	0.086*	0.243*	-0.166*	-0.325*	-0.064	-0.231*	1.000				
(15) Efficiency measures	-0.004	-0.037	-0.016	0.095*	-0.063	-0.054	0.074	-0.102*	0.050	-0.040	-0.079*	-0.016	-0.056	-0.095*	1.000			
(16) SRI dummy	-0.009	-0.043	0.031	0.008	-0.018	0.022	0.002	0.099*	0.004	-0.054	-0.105*	-0.021	-0.074	-0.126*	-0.031	1.000		
(17) SPeriod 1981-2000	-0.051	-0.022	-0.256*	0.046	0.040	0.066	-0.042	0.359*	0.132*	-0.023	-0.121*	0.026	0.170*	0.141*	-0.080*	-0.019	1.000	
(18) SPeriod 2001-2010	0.078*	-0.051	-0.026	-0.006	-0.104*	0.047	-0.095*	-0.238*	0.072	0.076	-0.036	-0.009	-0.081*	-0.048	0.120*	-0.004	-0.636*	1.000
(19) Conditional evaluation	-0.020	-0.081*	-0.102*	0.156*	0.024	-0.139*	0.083*	0.160*	0.322*	0.060	-0.150*	-0.030	-0.072	0.360*	-0.044	-0.058	0.062	0.022
(20) Index benchmark	0.031	0.115*	0.028	-0.102*	0.140*	0.080*	-0.043	0.011	-0.119*	0.153*	0.114*	-0.015	-0.116*	-0.289*	0.101*	-0.024	-0.265*	0.156*
(21) One-to-one matching	-0.008	-0.055	-0.120*	0.105*	-0.006	0.013	-0.014	0.074	-0.133*	-0.045	0.027	0.153*	0.053	-0.055	-0.040	0.010	-0.006	0.095*
(22) One-to-many matching	-0.056	-0.073	-0.198*	0.024	-0.167*	0.217*	0.114*	0.217*	0.253*	0.038	-0.218*	0.031	0.088*	0.148*	-0.069	0.031	0.172*	0.022
(23) NumMatchCriteria	-0.038	-0.050	-0.201*	0.138*	-0.091*	0.092*	0.083*	0.239*	0.088*	-0.026	-0.124*	0.012	0.110*	0.095*	-0.071	0.064	0.108*	0.073
(24) Peer-reviewed	0.016	0.065	0.132*	-0.070	-0.029	0.068	0.395*	0.373*	0.010	0.020	-0.039	-0.012	0.020	-0.094*	0.060	0.039	-0.076*	-0.120*
(25) Stock	0.094*	0.198*	0.406*	-0.249*	0.122*	-0.143*	-0.179*	-0.186*	-0.272*	0.011	0.225*	-0.057	-0.149*	-0.139*	-0.084*	-0.038	-0.310*	0.085*
(26) Survivorship bias	-0.035	-0.100*	-0.290*	0.183*	-0.070	0.112*	0.202*	0.248*	0.224*	-0.051	-0.333*	0.018	0.124*	0.273*	-0.126*	0.100*	0.374*	-0.134*
(27) Bloomberg	-0.032	-0.027	0.154*	-0.047	0.051	-0.060	0.053	0.003	-0.060	-0.052	-0.047	0.013	0.084*	0.036	0.179*	-0.044	0.033	-0.010
(28) CRSP	-0.060	0.030	-0.160*	-0.256*	-0.125*	0.439*	-0.017	0.189*	-0.020	-0.051	-0.148*	0.025	0.105*	0.099*	-0.047	0.101*	0.270*	-0.072
(29) Morningstar	-0.012	-0.028	-0.077*	-0.002	0.007	0.116*	0.362*	-0.023	0.129*	0.035	-0.040	0.068	-0.014	-0.019	0.089*	-0.033	-0.092*	0.145*
(30) Thomson Reuters	0.010	-0.050	-0.136*	0.394*	-0.182*	-0.192*	-0.044	0.044	0.119*	0.105*	-0.064	-0.084*	-0.017	0.100*	-0.123*	-0.021	0.107*	-0.012
(31) ESG taxonomy	-0.044	-0.029	-0.379*	0.155*	-0.046	0.188*	0.139*	0.051	0.310*	0.042	-0.217*	0.053	0.034	0.095*	-0.022	0.099*	0.180*	-0.012
(32) Bloomberg taxonomy	-0.088*	0.066	-0.072	-0.105*	0.201*	0.038	0.105*	-0.119*	-0.049	-0.022	0.065	-0.021	-0.003	-0.074	0.022	-0.041	-0.052	-0.082*
(33) Morningstar taxonomy	-0.022	-0.023	-0.195*	0.080*	-0.084*	0.170*	0.217*	0.206*	0.106*	0.020	-0.111*	0.037	0.008	0.056	-0.065	0.021	-0.010	-0.088*
(34) SIF taxonomy	-0.011	0.030	-0.170*	-0.283*	-0.077*	0.428*	-0.099*	-0.041	-0.051	-0.062	-0.037	0.097*	0.089*	0.033	-0.056	0.020	0.037	0.090*
(35) SocialFunds taxonomy	-0.006	-0.066	-0.078*	-0.130*	-0.071	0.223*	-0.021	-0.037	0.177*	-0.046	-0.001	-0.018	-0.036	0.074	-0.026	-0.035	-0.096*	0.163*
(36) Vigeo taxonomy	-0.009	0.002	-0.124*	0.272*	-0.113*	-0.136*	0.011	0.150*	0.172*	-0.001	-0.113*	-0.028	0.045	0.027	0.198*	-0.024	0.008	0.110*

Variables	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)
(1) Performance difference																		
(2) Outperformance of SRI																		
(3) Asia																		
(4) Europe																		
(5) Rest of the World																		
(6) US																		
(7) Impact factor																		
(8) Finance journal																		
(9) number of publications																		
(10) Excess returns																		
(11) Risk-adjusted measures																		
(12) Other returns																		
(13) One-factor model																		
(14) Multi-factor model alpha																		
(15) Efficiency measures																		
(16) SRI dummy																		
(17) SPeriod 1981-2000																		
(18) SPeriod 2001-2010																		
(19) Conditional evaluation	1.000																	
(20) Index benchmark	-0.088*	1.000																
(21) One-to-one matching	-0.076	-0.132*	1.000															
(22) One-to-many matching	0.122*	-0.209*	-0.120*	1.000														
(23) NumMatchCriteria	0.068	-0.229*	0.570*	0.574*	1.000													
(24) Peer-reviewed	0.107*	0.058	-0.041	0.067	0.038	1.000												
(25) Stock	-0.159*	0.220*	-0.145*	-0.205*	-0.214*	-0.088*	1.000											
(26) Survivorship bias	0.174*	-0.244*	-0.009	0.291*	0.201*	-0.080*	-0.353*	1.000										
(27) Bloomberg	0.049	0.011	-0.010	-0.118*	-0.073	0.098*	-0.003	-0.017	1.000									
(28) CRSP	-0.024	-0.128*	0.060	0.236*	0.215*	-0.012	-0.121*	0.370*	0.184*	1.000								
(29) Morningstar	-0.060	-0.135*	0.038	0.121*	0.023	0.221*	-0.256*	0.032	-0.146*	-0.224*	1.000							
(30) Thomson Reuters	0.147*	-0.025	0.074	0.051	0.150*	-0.112*	-0.056	0.137*	-0.139*	-0.228*	-0.321*	1.000						
(31) ESG taxonomy	0.025	0.030	-0.025	0.304*	0.185*	-0.080*	-0.202*	0.367*	-0.044	0.351*	0.045	0.073	1.000					
(32) Bloomberg taxonomy	-0.058	0.150*	-0.053	0.010	-0.056	0.112*	0.109*	-0.008	0.007	0.120*	-0.073	-0.164*	0.163*	1.000				
(33) Morningstar taxonomy	-0.029	-0.018	0.023	0.321*	0.156*	0.228*	-0.167*	0.119*	-0.147*	0.101*	0.269*	-0.076*	0.346*	-0.086*	1.000			
(34) SIF taxonomy	-0.073	-0.072	0.052	0.097*	0.111*	-0.129*	-0.205*	0.248*	-0.023	0.454*	0.008	-0.171*	0.301*	0.067	-0.058	1.000		
(35) SocialFunds. taxonomy	0.018	0.004	-0.045	0.256*	0.155*	-0.096*	-0.094*	0.105*	-0.059	0.032	-0.032	0.045	0.138*	-0.035	0.051	0.295*	1.000	
(36) Vigeo taxonomy	0.100*	-0.137*	0.073	0.299*	0.327*	0.012	-0.122*	0.069	-0.093*	0.050	0.029	0.178*	0.220*	-0.055	0.113*	-0.101*	-0.047	1.000

** shows significance at p<.05*

Table A2. Studies Included in the Analyses

Id	Authors	Publication	In Meta-Regression	SJR Finance journal	Peer-reviewed journal	2019 Clarivate Impact factor	Study Considers Stocks (Funds otherwise)
1	Humphrey and Lee	2011	1	0	1	4.141	0
2	Benson et al.	2006	1	0	1	4.141	0
3	Hamilton et al.	1993	1	1	1	1.682	0
4	Koellner et al.	2007	1	0	1	6.539	0
5	Bauer et al.	2006	1	1	1	2.382	0
6	Gregory et al.	1997	1	1	1	1.473	0
7	Kreander et al.	2005	1	1	1	1.473	0
8	Bauer et al.	2005	1	1	1	2.269	0
9	Ayadi et al.	2016	1	1	1	1.587	0
10	Bollen	2007	1	1	1	2.707	0
11	Gregory and Whittaker	2007	1	1	1	1.473	0
12	Chang and Witte	2010	1	0	1	0	0
13	Sánchez and Sotorrío	2009	1	0	0	0	0
14	Van Liedekerke et al.	2007	1	0	0	0	0
15	Derwall and Koedijk	2009	0				
16	Bello	2005	1	1	1	1.263	0
17	Statman	2000	1	1	1	1.682	0
18	Kempf and Osthoff	2008	1	1	1	1.473	0
19	Bauer et al.	2007	1	0	1	4.141	0
20	Mueller	1991	1	0	1	0	0
21	Goldreyer and Diltz	1999	1	1	1	0	0
22	Gil-Bazo et al.	2010	1	0	1	4.141	0
23	Renneboog et al.	2008	1	1	1	2.521	0
24	Rubio et al.	2018	1	0	1	1.317	0
26	Reddy et al.	2017	1	0	1	1.93	0
27	Basso and Funari	2014	1	0	1	4.213	0
28	Hwang et al.	2011	1	0	1	0	0
29	Basso and Funari	2014	1	0	1	2.987	0
30	Alda	2017	1	0	1	0	0
31	Matallín-Sáez et al.	2019	1	0	1	7.246	0

32	Climent and Soriano	2011	1	0	1	4.141	0
34	Renneboog et al.	2011	1	1	1	2.82	0
36	Gasser et al.	2017	1	0	1	4.213	1
38	Brzeszczyński and McIntosh	2014	1	0	1	4.141	1
39	Ito et al.	2013	1	0	1	2.175	0
40	Bodhanwala and Bodhanwala	2019	1	0	1	0	1
41	Chang et al.	2019	0				
42	Kadiyala, P.	2009	1	1	1	0	0
43	Das and Rao	2013	0				
44	Nofsinger and Varma	2014	1	1	1	2.269	0
45	Filbeck et al.	2016	1	0	1	0	0
46	Leite and Cortez	2014	1	1	1	1.801	0
47	Rahman et al.	2017	1	1	1	0	0
48	Renneboog et al.	2012	0				
49	Utz and Wimmer	2014	1	0	1	0	0
50	Munoz et al.	2014	1	0	1	4.141	0
51	Belghitar et al.	2017	1	1	1	1.263	0
52	Ng and Zheng	2018	1	0	1	5.203	1
53	Ielasi and Rossolini.	2019	0				
55	Drut	2010	0				
56	Capelle-Blancard and Monjon	2014	0				
57	Day et al.	2016	1	0	1	0	0
58	Cortez et al.	2008	1	0	1	4.141	0
59	Rodríguez	2010	1	1	1	0	0
60	Silva and Cortez	2016	1	0	1	7.246	0
61	Baeza-Sampere et al.	2016	1	1	1	0.48	0
63	López-Arceiz et al.	2018	1	0	1	4.141	0
66	Alvarez and Rodríguez	2015	1	0	1	0	0
67	Hachenberg and Schiereck	2018	0				
70	Tripathi and Bhandari	2015	1	0	1	0	1
71	Henke	2016	0				
72	Lööf and Stephan	2019	0				
73	Gómez-Bezares et al.	2016	1	0	1	2.576	1

76	Jin et al.	2006	1	1	1	2.382	0
77	Lee et al.	2010	1	1	1	2.217	0
78	Mollet and Ziegler	2012	1	1	1	0	1
79	Borgers et al.	2015	1	1	1	2.269	0
80	Chung et al.	2012	1	1	1	0	0
81	Chang et al.	2012	1	0	1	0	0
82	Ibikunle and Steffen	2017	1	0	1	4.141	0
87	Patel	2016	1	0	0	0	0
89	Srivastava	2017	0				
90	Alda and Vicente	2020	1	1	1	1.801	0
91	Jin and Han	2018	0				
92	Halkos and Sepetis	2007	1	0	1	4.482	1
93	Marco et al.	2011	1	0	1	0	0
94	Reboredo et al.	2017	1	0	1	12.11	0
96	Qoyum et al.	2020	1	1	1	2.13	1
97	Abdelsalam et al.	2014	0				
99	Azmi et al.	2020	1	1	1	2.382	0
101	Munoz et al.	2014	0				
102	Allevi et al.	2019	1	0	1	5.203	0
104	Utz	2017	1	1	1	0	1
105	Abdelsalam et al.	2014	0				
106	Leite and Cortez	2015	1	1	1	2.497	0
107	Syed	2017	1	1	1	1.801	0
108	Sauer	1997	1	1	1	0	0
110	Lam et al.	2015	0				
114	Munoz	2016	0				
115	Alda	2019	1	1	1	3.527	0
116	Chang et al.	2019	0				
120	Lesser et al.	2015	0				
121	Marti-Ballester	2019	1	0	1	6.274	0
122	Gianfrate and Peri	2019	0				
130	Yesuf and Aassouli	2020	1	0	1	0	0
132	Chong and Phillips	2016	1	0	1	0	0
133	Goyal and Aggarwal	2014	1	0	1	0	1
134	Plagge and Grim	2020	0				
135	Madhavan and Sobczyk	2020	0				
136	Filbeck et al.	2019	1	1	1	0	1
137	Chen and Scholtens	2016	0				

138	Yue et al.	2020	1	0	1	2.576	0
139	Wee et al.	2020	1	0	1	0	0
140	Kiymaz	2019	1	0	1	0	0
143	Leite and Cortez	2014	1	0	1	0.752	0
144	Leite and Cortez	2018	1	1	1	0.974	0
145	Alda	2016	1	1	1	0	0
146	Noman et al.	2019	0				
147	Ang and Lean	2013	1	0	0	0	0
148	Larsen	2013	0				
149	Angelica and Utama	2020	1	0	1	0	1
150	Jones et al.	2008	0				
151	Zhang	2014	1	0	1	0	0
152	Basso and Funari	2005	1	0	0	0	0
153	Thomson	2013	1	0	0	0	0
155	Trinks et al.	2017	1	0	0	0	1
156	Sánchez and Sotorrío	2014	1	0	1	1.275	0
157	Mill	2006	1	0	1	4.141	0
158	Derwall	2007	1	0	0	0	0
159	Maroof and Javid	2016	0				
160	Schröder	2004	0				
162	Costa et al.	2011	1	1	1	0	0
163	Barwick-Barrett	2015	1	0	0	0	0
165	Blanchett	2010	0				
166	Elaut et al.	2015	1	0	0	0	1
167	Geczy et al.	2005	1	0	0	0	0
168	Wei	2018	1	0	0	0	0
169	Ferruz et al.	2012	1	0	1	4.141	0
170	Ang and Lean	2013	1	0	0	0	0
171	Mackie et al.	2018	1	0	0	0	0
172	Munoz et al.	2015	1	1	1	1.491	0
173	Amenc and Le Sourd	2010	1	0	0	0	0
174	Dolvin et al.	2019	0				
175	Ben Slimane et al.	2019	0				
176	Brzeszczyński et al.	2019	1	0	1	2.394	1
177	Jarno	2017	0				
178	Smimou and Ayadi	2019	1	0	1	0	0
179	Lundberg et al.	2009	1	0	0	0	0
180	Hunt and Weber	2019	1	0	1	3.333	1
181	Castro	2015	1	0	0	0	0
182	Adrianto	2016	0				

183	Chelawat and Trivedi	2013	1	0	1	0	1
185	Auer	2016	1	0	1	4.141	1
186	Hoepner and Nilsson	2018	0				
187	Marti-Ballester	2020	1	0	1	5.483	0
191	Arbelaez et al.	2006	1	0	0	0	1
192	Hoepner and Yu	2010	0				
194	Ferruz et al.	2012	1	0	0	0	0
195	Gueckel	2017	1	0	0	0	1
197	Deshmukh	2012	1	0	0	0	0
198	Bhatt et al.	2014	1	0	0	0	1
199	Tripathi and Jham	2020	1	0	1	0	1
200	Rompotis	2016	1	1	1	0	0