Asia-Pacific Journal of Financial Studies

Asia-Pacific Journal of Financial Studies (2019) **00**, 1–35 doi:10.1111/ajfs.12244

Can Socially Responsible Investments Be Compatible with Financial Performance? A Meta-analysis*

Chang-Soo Kim**
College of Government and Business, Yonsei University-Wonju, Republic of Korea

Received 29 March 2018; Accepted 22 November 2018

Abstract

This paper examines whether socially responsible investment (SRI) outperforms conventional investments by meta-analysis. The result shows that weighted average effect size is not significantly different from zero, suggesting that SRI performance is not different from conventional investments. Meta-regression results regarding sampling issues indicate that economic crisis, control group, investment universe, screening procedure, and mutual funds are important determinants of effect sizes, and publication year, author type, and control group are important factors of the absolute value of effect sizes. Regarding methodological issues, risk adjustment, weighting scheme, data refinement, benchmark model, and matching procedure are significant factors determining the absolute value of effect sizes.

Keywords Effect size; Meta-analysis; Meta-regression; Random-effect model; Socially responsible investment

JEL Classification: G11, G32, M14

1. Introduction

Researchers and investors have long been curious about whether socially responsible investment (SRI) funds and portfolios financially outperform conventional ones. Early studies date back to the 1970s (Moskowitz, 1972; Alexander and

*This paper was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5B8925203). This work was supported by Fulbright Senior Researcher Scholarship.
**Corresponding author: Department of Business Administration, Yonsei University, Yonsei-dae-gil 1, Heungup-myon, Wonju, Gangwon-do 26493, Republic of Korea. Tel: +82-33-760-2331, Fax: +82-33-760-2918, email: kimc@yonsei.ac.kr.

Buchholz, 1978), and a significant amount of research has been accumulated so far.[1] However, existing evidence shows mixed results. For example, Auer (2016) shows that investors can do financially well while doing social good based on environmental, social, and governance ratings data for the European market. However, Renneboog *et al.* (2008a) report that the SRI funds of France, Ireland, Sweden, and Japan significantly underperform conventional funds by 4–7% per annum during the period 1991–2003. In addition, there are many other papers that report no difference in performance between SRI and conventional funds (Hamilton *et al.*, 1993; Kreander *et al.*, 2005; Kempf and Osthoff, 2008; Hoepner and Schopohl, 2018).

It is imperative to determine the overall influence of SRI on investment performance, because if the impact is positive, investors will voluntarily participate in socially beneficial investments and government does not have to intervene in the market to promote social goals. If the impact is negative, however, investors have to sacrifice their investment benefits to pursue their social preferences and the public sector has to intervene in one way or another to encourage SRI activities. The direction and magnitude of SRI influence can also have a profound implication for financial industries with respect to product development, asset management, fiduciary duty, performance evaluation, and so forth. Understanding the impact of SRI is also important to academic research. As Combs *et al.* (2011) point out, without a proper synthesis of existing literature more fruitful further research is difficult.

Despite this, there have been few attempts to synthesize existing literature on SRI and even these have not been successful. For example, Rathner (2013) lacks representativeness due to a small sample size and incomplete methodology. Revelli and Viviani (2015) performed a meta-analysis based on 85 studies. However, they fail to draw a sharp conclusion due to ambiguity caused by pooling several countries in one sample. López-Arceiz *et al.* (2018) show that culture is an important determinant of SRI influence, so combining countries with different cultural backgrounds will prevent researchers from obtaining a correct picture of SRI.[2] More importantly, they analyze the impact of moderators only on effect size (hereafter, ES) ignoring the impact on the absolute value of effect size (hereafter, AES),

[1]SRI is different from corporate social responsibility (CSR). CSR focuses on corporate social activities and papers on CSR examine whether firms can improve financial performance by engaging in socially desirable activities. SRI focuses on investment strategies that employ various screening procedures (e.g., positive, negative, and best-in-the-class) based on so-called ESG (environmental, social, and governance) criteria and papers on SRI compare the investment performance of these portfolio strategies with conventional ones.

[2]Moreover, their study seems to have multiple errors. For example, they include Smith (1996), which is an event study, even though they explicitly mention that they do not include event studies in the text. Gregory *et al.* (1997) have a Hedges' g value of 0.24 in their Appendix, but the sign should be negative since the ethical funds performed worse than the non-ethical funds.

although many valuable conjectures can be made regarding AES. This point is especially important in the current situation where positive and negative effects coexist, since focusing only on the level of ESs will average out positive and negative impacts leading researchers to underestimate the impact of SRI. To our knowledge, our paper is the first that investigates the influence of moderators on both ESs and AESs.

Another reason for the lack of sound summary papers is the diversity and complexity of existing studies with regard to samples, methodologies, performance measures, investment universe, benchmarks, etc. A meta-analysis is a very useful scientific tool to synthesize a wide variety of heterogeneous papers. It calculates ES that can be interpreted as a standardized difference between performances of SRI and conventional portfolios. By standardizing performance differences, whatever the performance measures, a common measure of difference between the two types of portfolios is generated, making synthesis possible across very diverse papers.

Given the situation in SRI research, this article attempts to achieve the following goals. First, we provide a sound synthesis of existing literature which shows an overall influence of SRI on investment performance in terms of sign and magnitude through a meta-analysis. Then, we examine a variety of determinants of SRI influence, since it is conceivable that the magnitude and direction of SRI effects may have a systematic relationship with sample characteristics and research methodologies. The analyses on moderators will help us to produce interesting implications for future research and practice. When performing the analysis, we carefully explain our meta-analysis and data collection process to ensure replicability.

This study is organized as follows. In the next section, we develop hypotheses regarding the overall impact of SRI and moderators that can have a systematic relationship with ESs and AESs. In Section 3, we explain the meta-analysis and the concept of ES and show definitions of moderators used for hypothesis testing. In Section 4, we present our sampling process in detail, since sampling is crucial in performing and interpreting a meta-analysis. Section 5 reports the meta-analysis results, and Section 6 presents conclusions and discussions.

2. Hypotheses

Our initial research interest is on the average impact of SRI on investment performance. Thus, we first propose a hypothesis on the overall ES and then proceed to establish hypotheses for moderators that may have a systematic relationship with ESs and AESs. We attempt to employ moderators to reflect research characteristics as extensively as possible based on theory and practice. We split the moderators into two broad categories: sampling and methodology. Sampling relates to data collection, and methodology is about analysis tools.

2.1. Overall Effect Size

The CSR literature shows that CSR firms perform better than non-CSR firms (Wu, 2006; Margolis *et al.*, 2009). Since SRI invests more in CSR firms, the performance of SRI will be higher than conventional investments. SRI performance can also be better since SRI fund managers may have a deeper knowledge of their investment pools (Girard *et al.*, 2007; Gil-Bazo *et al.*, 2010). SRI screening eliminates unethical firms from the investment universe, and fund managers can focus on a relatively small number of firms. In fact, Kacperczyk *et al.* (2005) show that more concentrated portfolios produce a higher return than broadly diversified portfolios. SRI fund turnover is lower since SRI investors are more loyal to social values than conventional investors, and the resulting lower transaction costs also predict a better performance.

On the other hand, portfolio theory predicts an inferior performance of SRI portfolios. Other things being equal, SRI constraints on the investment universe will lead to a less favorable efficient frontier, sacrificing diversification benefits (Girard *et al.*, 2007; Renneboog *et al.*, 2008a). For example, excluding sin stocks can cause investors to sacrifice returns, since they generate a higher risk-adjusted return than conventional stocks (Fabozzi *et al.*, 2008; Hong and Kacperczyk, 2009; Durand *et al.*, 2013; Borgers *et al.*, 2015). The higher transaction costs and management fees of SRI funds also predict an inferior performance, since they are relatively small, losing economies of scale (Bauer *et al.*, 2005; Barnett and Salomon, 2006). Ayadi *et al.* (2011) find that the average expense ratio of Canadian SRI fund is 1.97%, which is much larger than the 1.29% global benchmark for conventional funds.

Finally, any impact of SRI cannot be justified theoretically. The efficient market hypothesis predicts that SRI portfolios and funds cannot produce abnormal returns, since investors and fund managers can easily replicate ethical screening procedures which are public information. Even though some SRI funds seem to outperform conventional funds, their performance superiority may vanish if transaction costs are subtracted.

On the empirical side, evidence is also mixed with positive (Hickman and Teets, 1999; Guenster *et al.*, 2005; Brander, 2006; Kempf and Osthoff, 2007; Galema *et al.*, 2008; Peiris and Evans, 2010) and negative (Mueller, 1991; Girard *et al.*, 2007; Durand *et al.*, 2013) impacts for SRI. In addition, there are many papers reporting no impact. For example, Kreander *et al.* (2005) uncover similar performances between SRI and conventional funds when they investigate the SRI funds of seven European countries. Bauer *et al.* (2007) examine a sample of Canadian ethical funds and again find no significant difference between SRI and conventional funds. In addition, there are still many papers that report no difference between SRI and conventional investments (Sauer, 1997; Hutton *et al.*, 1998; Benson *et al.*, 2006; Kempf and Osthoff, 2008; Renneboog *et al.*, 2008a; Humphrey *et al.*, 2016). Therefore, we establish a null hypothesis on the overall effect of SRI as follows.

Hypothesis 1. The influence of SRI will not be different from zero.

2.2. Sampling Issues

2.2.1. Time.

Among various sampling issues that generate the SRI performance documented in many studies, the first-order effect could be caused by the change in the conceptualization of social responsibility. Over the past three decades, not only firms but also investors have changed their views on CSR. In earlier years, the CSR strategy may have been adopted by innovative firms and the SRI strategy, including these firms, may have outperformed conventional investment strategies. However, the relative advantage of SRI may have weakened over time since the uniqueness of SRI has decreased as more firms promote CSR activities. This process will lead to a lower abnormal return over time. It is also possible that investing communities will learn from academic studies. As more investors use SRI investment strategy via the learning process, the return can diminish. Consistent with this argument, Hoepner and Schopohl (2018) show that estimates for the impact of SRI become marginally significant or lose their significance over the sub-period 2008–2015 when compared with the entire sample period 2001–2015.

However, the dampening impact SRI process may not be a uniformly decreasing one, but rather alternate in sign. Research in general tends to show three dialectical stages of development (O'Connor, 2003). A thesis is posited in the beginning and it gives rise to its reaction, an antithesis, which contradicts or negates the thesis. Finally, a synthesis resolves the tension between the two. We can expect that research in the SRI area may also follow a similar pattern. Since the merits of SRI were overemphasized in the beginning, an argument of negative impact shows up as an antithesis which is followed by a synthesis that supports no or less impact. As for the degree of fluctuation, it will decrease over time. Since methodologies have become more sophisticated over time, any spurious difference in performance between SRI and conventional portfolios caused by misspecification of methodologies will become smaller over time. Thus, we establish the following hypothesis.

Hypothesis 2-1. ESs are positive in the beginning and change signs with decreasing absolute value.

During times of economic crisis, portfolio performance reduces compared to normal periods. The SRI performance will follow a similar pattern and the ESs computed from the sample period including a crisis will be different from those in normal periods. Thus, we establish the following hypothesis.

Hypothesis 2-2. Papers whose sample periods include a crisis period will show smaller ESs than papers whose sample periods do not include a crisis.

2.2.2. Publication.

One of the important considerations in meta-analysis regarding sampling is publication bias. As the phrase "Publish or Perish" signifies, publication is imperative in academia (Zivney and Bertin, 1992). Since so many papers are submitted, academic journals have become very selective and papers with a high statistical significance are more likely to be published than papers with insignificant results (Song *et al.*, 2000). According to Stanley (2005), we can suspect that type II selection error, that is, a tendency to report significant results excessively, exists if the values of standardized ESs exceed 1.96 more than 5% of the time. Other things being equal, if a publication bias exists, published papers will have higher ESs, either positive or negative. In order to check whether our sample has publication bias, we establish the following hypothesis.

Hypothesis 3. The AESs for published papers will be higher than those for unpublished papers.

2.2.3. Author type.

Papers written by professional managers and scholars can show differences in ESs since they are quite different in various respects, including incentives for performing research. For example, Vaihekoski (2004) mentions that academic researchers are, in general, concerned with aspects related to asset pricing, whereas institutional investors are primarily interested in detecting opportunities that can generate superior risk-adjusted performance. Asset pricing is closely related to the efficient market hypothesis, and insignificant differences between SRI performances and conventional portfolios can be accepted as confirming market efficiency. However, if a large difference is found between the two performances, justification regarding the causes of the difference is often quite challenging. Among industry professionals, however, market efficiency is not accepted as a norm, so researchers from financial industries may pay more attention to finding profit potential than supporting market efficiency. Therefore, we conjecture that researchers in academia tend to find a smaller performance difference between SRI and conventional groups than industry professionals. In order to test this argument, we propose the following hypothesis.

Hypothesis 4-1. The AESs obtained by academic researchers will be smaller than those obtained by industry researchers.

It is possible that researchers in different disciplines may have different perspectives on SRI and this will lead to a systematic difference in their research outcomes. For example, researchers in the finance area may focus more on profit maximization and argue against the benefit of SRI, while researchers in other areas such as business ethics may think social benefits are more important than monetary benefits, advocating for SRI. To examine the impact of a difference in researcher perspectives on ESs, we establish the following hypothesis.

Hypothesis 4-2. The ESs of papers published in finance journals will be smaller than those published in non-finance journals.

2.2.4. Control group.

Researchers compare the performance of SRI funds and portfolios with the performance of various control groups. For example, Muñoz *et al.* (2015) compare SRI mutual funds with conventional mutual funds, while Edmans (2011) compares SRI portfolios with market indexes. In general, the difference between SRI portfolios and market indexes is higher than that between SRI portfolios and corresponding conventional portfolios. Whereas SRI portfolios suffer from diversification loss due to screening procedures, market indexes do not. On the other hand, conventional portfolios are formed by an author(s) considering comparability with SRI portfolios. Thus, we propose the following hypothesis.

Hypothesis 5-1. The comparison between SRI portfolios and market indexes will have larger AESs than the comparison between SRI and conventional portfolios

For the papers that use conventional market indexes as the control group, the treatment groups are either SRI portfolios or SRI indexes (Statman, 2006). In this case, the discrepancy between the two indexes, i.e., SRI and conventional, is smaller than that between SRI portfolios and conventional indexes. For example, the Domini Social Index is a leading market capitalization-weighted SRI index. It was created in 1990 by starting with the companies in the S&P 500 index, so many companies are included in both indexes (Di Bartolomeo and Kurtz, 1999; Brander, 2006). Therefore, we propose the following hypothesis.

Hypothesis 5-2. The AESs will be smaller when SRI indexes rather than SRI portfolios are compared with conventional indexes.

Socially responsible investment portfolios have a less favorable risk–return trade-off due to diversification loss (Hong and Kacperczyk, 2009; Statman and Glushkov, 2009). When market portfolios are used rather than conventional portfolios composed by researchers as a control group, the degree of diversification loss is relatively larger. This leads to the following hypothesis regarding ESs themselves.

Hypothesis 5-3. Experiments using market index as a control group will have smaller ESs than those using conventional portfolios as a control group.

2.2.5. Investment universe.

Cultural elements are important in studying the impact of SRI (Hill *et al.*, 2007; López-Arceiz *et al.*, 2018). Since fund managers will adopt investment strategies that incorporate values appreciated in their culture (Louche and Lydenberg, 2006), SRI funds can appeal more strongly to regions with a similar cultural background than to other countries with different cultural backgrounds. Indeed, Jones *et al.* (2008) find that SRI funds with a high exposure to international capital markets underperform the market by 3% per annum. Therefore, we establish the following hypothesis.

Hypothesis 6. The performance of domestic SRI funds will be better than international SRI funds.

2.2.6. Screening procedure.

Funds using negative screening exclude unethical firms, and sometimes even entire industries, from their portfolios, causing a sharp shrink in the portfolio frontier. Moreover, since sin stocks seem to perform better, SRI based on negative screening will result in a worse performance than conventional investment (Fabozzi *et al.*, 2008; Hong and Kacperczyk, 2009). This problem is less severe in the case of positive screening, since fund managers can add SR stocks to their existing portfolios. Even when SRI managers select stocks purely based on positive criteria, they may perform better than broadly diversifying fund managers since they may have an intimate knowledge of SR stocks (Kacperczyk *et al.*, 2005; Girard *et al.*, 2007). Based on these arguments, we establish the following hypothesis.

Hypothesis 7. Funds employing a positive screening procedure will outperform those with a negative screening procedure.

2.2.7. Mutual funds.

In existing literature, many authors study the performance of SRI mutual funds (Geczy *et al.*, 2005; Chang and Witte, 2010; Climent and Soriano, 2011; Muñoz *et al.*, 2014), while others study SRI portfolios formed by themselves (Diltz, 1995; Gompers *et al.*, 2003; Core *et al.*, 2006). There can be differences between the two types of research since it is well known that managed funds rarely beat the market (Jensen, 1968; Ferson and Schadt, 1996; Ferson and Warther, 1996; Carhart, 1997). In addition, funds formed by researchers do not consider market frictions such as liquidity constraints and transaction costs explicitly, which is not the case in the analysis of mutual funds (Adamsson and Hoepner, 2015). Based on these arguments, we establish the following hypothesis.

Hypothesis 8. ESs from examining SRI mutual funds will be lower than ESs from examining portfolios formed by researchers.

2.3. Methodological Issues

Effect sizes can also vary with methodologies employed.[3] In particular, under efficient capital markets, there should not be a performance difference between SRI and conventional portfolios if all relevant factors are considered. However, if some important determinants are omitted by methodological misspecification, it will

---

[3]Although methodological issues exist for most asset pricing/capital market studies, they are particularly important in the meta-analysis of SRI. The result on SRI performance can be systematically different depending on the research methodologies used. Thus, investigating the impact of methodological difference on ESs from a variety of different angles can significantly enhance our understanding of SRI.

produce a spurious difference in portfolio performances. For instance, Gregory *et al.* (1997) reveal that the outperformance of UK SRI funds found in previous studies disappears when they improve methodology by using a two-factor benchmark to control size. In general, the more rigorous the methodologies, the smaller the possible confounding effects.

Existing literature examines various methods of controlling for these problems, and this paper categorizes these into three groups, that is, variable adjustment (e.g. risk adjustment, value-weighting, splits and dividends adjustment), refining benchmark model (e.g. multi-factor model, conditional model), and matching.

2.3.1. Variable adjustment.

Performance proxies in the existing literature can broadly be categorized as measures with and without risk adjustment. Raw returns are the performance measure without considering risk, while Jensen's alpha, the Sharpe ratio, and the Treynor ratio are risk-adjusted measures. If raw returns are used even though the portfolios of SR and conventional firms have different risk profiles, there can be a spurious performance difference between SRI and conventional investment because they do not reflect proper risk factors (Becchetti and Ciciretti, 2009). Therefore, we establish the following hypothesis.

Hypothesis 9-1. The AESs will be smaller when risk-adjusted performance measures are used.

When calculating portfolio returns, value-weighting is theoretically more appropriate than equal-weighting. Value-weighting schemes are also in accord with typical financial industry practice (Humphrey and Tan, 2014), and most feasible for institutional investors who constitute the largest proportion of the SRI market (Trinks and Scholtens, 2017). In some cases, results based on value-weighting are different from those based on equal-weighting. For example, Statman and Glushkov (2009) find that excess returns of value-weighted portfolios are lower than those of equally weighted portfolios. If SRI mutual funds have a higher exposure to small firms than conventional funds (Luther and Matatko, 1994; Schröder, 2003), equal-weighting schemes can generate undeserved performance differences between SRI and conventional portfolios. In order to examine this effect, we establish the following hypothesis.

Hypothesis 9-2. The AESs calculated using a value-weighting scheme will be smaller than those using an equal-weighting scheme.

When calculating returns, some papers pay more attention to data accuracy by taking dividends and stock splits into account (Shank *et al.*, 2005), whereas other papers ignore them. Inaccurate data can cause a spurious difference between SRI and conventional portfolios. For example, if SRI funds in general invest more in small firms than conventional funds (Luther *et al.*, 1992; Luther and Matatko, 1994), ignoring dividends may produce unwarranted differences in the performance of SRI and conventional portfolios. This is because, other things being equal, the SRI universe would have a lower dividend payout ratio than the conventional funds universe due to the higher investment opportunities of small firms over large firms. Therefore, we establish the following hypothesis.

Hypothesis 9-3. The ESs will be smaller when returns are calculated reflecting stock splits and dividends.

2.3.2. Refining benchmark model.

Several authors examine the impact of SRI in the US market using a single-index framework (Hamilton *et al.*, 1993; Reyes and Grieb, 1998; Goldreyer and Diltz, 1999; Statman, 2000; Bello, 2005). However, the single-index model can produce distorted results by not controlling factors other than market portfolio. Prior literature indicates that SRI and conventional funds are different in their investment universe and portfolio strategies, having significantly different loadings on size, book-to-market, and momentum factors (Bauer *et al.*, 2005; Renneboog *et al.*, 2008b). In order to take into account the difference between the two types of funds, researchers use multi-factor models (Fama and French, 1993; Carhart, 1997; Galema *et al.*, 2008; Edmans, 2011). A spurious performance difference between SRI and conventional funds may disappear if all relevant factors are incorporated.

CAPM, the Fama-French model, and the four-factor Carhart model are unconditional models since coefficients of factors do not change over time. In practice, however, fund managers change their portfolio holdings according to changes in macroeconomic environments, resulting in dynamic risk–return profiles (Renneboog *et al.*, 2008a). In this situation, the unconditional model may be inappropriate, hence Ferson and Schadt (1996) and Christopherson *et al.* (1998) propose conditional models that allow coefficients of factors to vary over time. Since investors on SRI funds may not be as sensitive as investors on conventional funds to economic environment changes, SRI funds may not adjust their portfolio holdings as quickly as conventional funds (Bollen, 2007; Benson and Humphrey, 2008). If this behavioral difference is ignored in deriving performance measures by employing the unconditional model as a benchmark, SRI performance would become spuriously different from conventional performance. Thus, we propose the following hypothesis.

Hypothesis 10. The AESs calculated using more refined return-generating models will be smaller.

2.3.3. Matching.

Some authors employ a matching procedure where each SRI fund is mapped with conventional funds of similar characteristics like size, location, composition, and time horizon so as to control confounding factors (Derwall and Koedijk, 2009; Ferruz *et al.*, 2012; López-Arceiz *et al.*, 2018). If proper counterfactuals to SRI portfolios are employed, the spurious difference between SRI and conventional performance can be eliminated. Therefore, the AESs will be smaller when matching procedure is employed.

The above discussion on methodological issues leads to the following hypothesis.

Hypothesis 11. The AESs will be smaller when more sophisticated methodologies are employed.

3. Methodology

3.1. Meta-analysis

Since positive, negative, and no effect views can all be justified theoretically, synthesizing empirical findings to obtain the true overall impact of SRI is very important to both academia and industry. A meta-analysis is a particularly useful tool for this purpose, since it reflects the magnitude of performance differences between SRI and conventional groups by calculating ESs of individual studies and reporting overall ES utilizing information on the precision of individual studies (Combs *et al.*, 2011). Specifically, the first step of meta-analysis is to calculate the ES for each experiment or paper using Hedges' g as follows (Hedges, 1981):[4]

$$g = \frac{\overline{P^T} - \overline{P^C}}{\sigma}, \tag{1}$$

where $\overline{P^T} \cdot (\overline{P^C})$ is the mean performance of portfolios included in the SRI (conventional) group and σ is the pooled within-group standard deviation which can be computed using the following formula:

$$\sigma = \sqrt{\frac{(N^T - 1)V^T + (N^C - 1)V^C}{(N^T + N^C - 2)}}, \tag{2}$$

where N^T (N^C) is the sample size of the SRI (conventional) group and V^T (V^C) is the variance of the SRI (conventional) group. Intuitively, the ES can be interpreted as a standardized performance difference between the two groups whatever the performance measures are and Table 1 shows the interpretation of ESs. The value of d = 0.01 has a very small effect, while d = 2.0 has a huge effect. Although we can typically compute ESs using the above equations, there are other cases where ESs are computed using different equations. The Appendix provides a detailed explanation on how to achieve ESs in special cases.

Once we obtain the ESs of each experiment or paper, the next step is to compute the weighted ES by considering all papers in the sample so as to obtain the overall SRI impact on portfolio performance as proposed by Hasselblad and Hedges

[4]There are in general multiple analyses in one paper. We use the terms "analysis," "study," "experiment" and "comparison" interchangeably.

Table 1 Interpretation of effect size

This table contains interpretations for magnitudes of effect size and their references. Originally Cohen (1988) suggested three groups of ESs, small, medium, and large, and the other three ESs have been appended by Sawilowsky (2009). A similar interpretation can be made in cases of negative effect sizes.

Effect size	*d*	References
Very small	0.01	Sawilowsky (2009)
Small	0.20	Cohen (1988)
Medium	0.50	Cohen (1988)
Large	0.80	Cohen (1988)
Very large	1.20	Sawilowsky (2009)
Huge	2.00	Sawilowsky (2009)

(1995). The basic idea of weighting is to give higher weights to studies with more precise ES estimates, and the reciprocal of variance for each comparison is used as weights. When aggregating ESs, we use either a fixed-effect model or a random-effect model, depending on the assumption for variability. Details are in the Appendix. In addition, we delve into the possibility that SRI impacts on financial performance may vary systematically according to moderators and thus perform meta-regression analysis (Lipsey, 2001).

3.2. Variable Definitions

Table 2 summarizes all moderators and variables including dependent variables, either ESs or AESs. We split moderators into two broad categories, one for sampling and the other for methodology.

3.2.1. Sampling.

In order to investigate sampling issues, we establish moderators for time, publication status, researcher background, control groups, investment universe, screening procedure, and mutual funds. Each moderator can have more than one variable to test a hypothesis, and variable definitions are shown in Table 2. For example, three variables are used to test publication status. Since publication bias is about statistical significance, we use two variables, Publication and Significance. In addition, we use a Quality variable to examine whether there is a systematic difference in AES according to journal quality. To examine the influence of control groups on ESs and AESs, we use two variables. SRP_CI is used to test *Hypotheses 5-1* and *5-3* and SRI_CI is used to test *Hypothesis 5-2*.

3.2.2. Methodology.

As for methodology, we investigate issues related to variable adjustment (risk adjustment, portfolio weighting scheme, data accuracy), adding independent variables (multi-factor model, conditional model) and a matching procedure. To see the impact of the matching procedure, we also use two variables, Matching and Match_N. If

Table 2 Summary of moderators and hypothesis tests

This table summarizes variable definitions to be used for hypothesis testing with predicted signs. It also provides a brief summary of the results from hypothesis tests. Moderators are grouped into one associated with sampling and the other related to methodologies. [1](+) positively significant; (−) negatively significant; 0 insignificant.

			Prediction		Results[1]		
	Moderator	Variables	AES	ES	AES	ES	Results on hypothesis
Sample	Time	Pub_year = the year a paper is published	(−)		(−)		2-1: ESs are positive in the beginning and change signs with decreasing absolute value over time
		Crisis = 1 if sample period includes crisis, 0 otherwise		(−)		(−)	2-2: ESs are smaller for papers whose sample periods include crisis
	Publication	Publication = 1 if published, and 0 if not published	(+)		0		3: There is no publication bias
		Significance = 1 if statistically significant, and 0 otherwise	(+)		(+)		
		Quality = 1 if published in above median ranking in JCR, and 0 otherwise	(−)		(−)		
	Author type	Academia = 1 if author(s) is from academia, and 0 if author(s) is from industry	(−)		(−)		4-1: Academic researchers tend to report a smaller absolute value of ESs than industry researchers
		Fin_Jnl = 1 if published in finance journal, and 0 if published in non-finance journal		(−)		0	4-2: There is no difference in ESs between finance and non-finance researchers
	Control group	SRP_CI = 1 if control sample is conventional market index, and 0 if control sample is portfolios or mutual funds	(+)		(+)		5-1: The absolute value of ESs when SRI samples are compared with market indexes is bigger than that when SRI samples are compared with conventional portfolios

Table 2 (*Continued*)

	Moderator	Variables	Prediction AES	Prediction ES	Results[1] AES	Results[1] ES	Results on hypothesis
		SRI_CI = 1 if both treatment and control groups are indexes, and 0 if treatment groups are not indexes, but control groups are indexes	(−)		(−)		5-2: The absolute value of ESs is smaller when SRI indexes, not SRI portfolios, are compared with conventional indexes
				(−)		(−)	5-3: Experiments using market index as a control group to SRI have smaller ESs
	Investment universe	Domestic = 1 if investment universe is domestic market and 0 if it is international market		(+)		(+)	6: The performance of domestic SRI funds is better than international SRI funds
	Screening procedure	Screening = 1 if a fund employs a positive screening and 0 if a fund employs a negative screening		(+)		(+)	7: Funds employing positive screening outperform those with negative screening
	Mutual funds	Mutual = 1 if a paper examines mutual fund sample and 0 for others		(−)		(−)	8: ESs derived from papers examining SRI mutual funds are lower than ESs from papers analyzing portfolios formed by researchers
Methodology	Variable adjustment	Risk_Adj = 1 if risk-adjusted performance measures are used, and 0 if raw returns are used	(−)		(−)		9-1: The absolute value of ESs calculated using risk-adjusted returns is smaller than that using risk-unadjusted raw returns
		Weighting = 1 if portfolio returns are calculated using value-weighting schemes, and 0 if equal-weighting is used	(−)		(−)		9-2: The absolute value of ESs calculated using a value-weighting scheme is smaller than that using an equal-weighting scheme

Table 2 (*Continued*)

Moderator	Variables	Prediction AES	Prediction ES	Results[1] AES	Results[1] ES	Results on hypothesis
	Accuracy = 1 if dividends or stock splits are explicitly incorporated to calculate returns, and 0 otherwise	(−)		(−)		9-3: The absolute value of ESs become smaller if returns are correctly calculated reflecting stock splits and dividends
Adding independent variables	Index = 1 if multi-factor models are used, and 0 if a single-index model is used	(−)		(−)		10: The absolute value of ESs calculated using multifactor models is smaller than that using a single-index model
	Conditional = 1 if performance measure is derived by conditional models, and 0 if performance measure is derived by unconditional models	(−)		0		The absolute value of ESs does not have any significant relationship with the use of conditional models
Matching	Matching = 1 if SRI portfolios are matched with conventional ones of similar characteristics, and 0 if no matching	(−)		(−)		11: The absolute value of ESs is smaller when matching procedure is employed
	Match_N = The number of control firms to be matched	(−)		(−)		

multiple firms are matched to an SRI sample, potential errors caused by the peculiarity of a matching firm in the case of a one-to-one match can be minimized. This will in turn reduce the spurious difference between SRI and control portfolios.

4. Data

Since representativeness is crucial in meta-analysis, we try to collect as many papers as possible including unpublished ones to avoid publication bias. We search databases such as EBSCO, JSTOR, Science Direct, Springer Link, Wiley-Blackwell, and SSRN, and utilize web search engines like Google Scholar. We also attempt to collect papers from many academic journals, as shown in Table A2 of the Appendix.

A detailed description of our data collection process is shown in Table A1 and a flowchart in the Appendix. We first compose keywords that are related to SRI as extensively as possible. By using the phrase "socially responsible investment," we obtain 1,555 articles that include at least one word in the phrase. Then, we check all articles one by one to sort out those that have the potential to be included in our final list of papers. By using the next phrase, "socially responsible investing," we have 1,226 results. Since the two phrases are quite similar, most observations are duplicates of the first search and we obtain only two additional articles that have the potential to be included in the final sample. The process continues with new phrases until we use "ethical investing," as shown in Table A1 of the Appendix. Since too many duplicates appear by that time, we begin to use the exact phrase "social investing" to narrow down the search results. We download 151 papers from this keyword search. Another important data collection process is to scrutinize the reference section of selected papers, especially summary or synthesis papers (Renneboog *et al.*, 2008b; Chegut *et al.*, 2011; Capelle-Blancard and Monjon, 2014).

After potential papers are collected, we read them more carefully to decide which papers are to be included in the final sample. Since the quality of data is crucial for an accurate assessment of SRI effects, we pay very careful attention to the data compilation process. The following explains some of the steps we have taken to preserve data integrity. Further details are in the Appendix.

1 We drop qualitative analyses.
2 If there are several versions of the same paper, we use the most recent paper to avoid overrepresentation bias. For example, Derwall *et al.* (2004) is a working paper version of Derwall *et al.* (2005), and we use the latter.
3 We include studies with matching firms or studies without matching firms but with market indices as benchmarks of SRI performance.
4 We include all relevant analyses based on important dimensions such as investment universe, sub-periods, conditional versus unconditional regression, matching versus non-matching, value weighted versus equal weighted, risk-adjusted versus risk-unadjusted, one-factor model versus multi-factor model, etc.

After this screening process, we obtain 51 papers, 18 from the keyword search and 33 from reference checking, which study the impact of SRI on financial performance in the US. López-Arceiz *et al.* (2018) assert that the influence of SRI varies significantly with a difference in cultural background. Focusing on the US market helps us sharpen the interpretation of results without the need to consider cultural differences. Figure S1 shows the flowchart of the selection process. The publication dates of the final sample run from 1978 to 2016.

From the selected papers, we extract 205 experiments to calculate ESs. In other words, our number of observations is the number of experiments, not papers, since each paper reports multiple analyses. For example, if a paper reports two tables of comparison between SRI and conventional funds, with one based on raw return and the other based on Jensen's alpha, we end up having two observations. Since a paper often reports results based on various subsamples, different experiments of the same paper can have different values for the number of firms, sample period, and all other relevant variables. We report detailed information including ESs on all 205 experiments in Table A2 of the Appendix.

5. Results

5.1. Results on Effect Sizes

Table 3 shows the descriptive statistics of ESs calculated from 205 comparisons. The median and mean ESs are 0.015 and 0.016 that can be interpreted as having very small ES according to Table 1. This result is consistent with many existing papers (Hamilton *et al.*, 1993; Reyes and Grieb, 1998; Goldreyer and Diltz, 1999; Statman, 2000; Humphrey *et al.*, 2016). They find no significant difference in the performance of SRI and conventional funds in the US market, even though they examine different samples and time periods. The distribution of ESs is not normal since skewness and kurtosis are significantly different from normal distribution.

The first result of interest in our meta-analysis is the overall value of ESs that allows us to judge the impact of SRI on investment performance. In order to calculate the weighted ES, we use the reciprocal of variance of each comparison. Table 4 shows weighted ESs from running fixed and random-effect models. The weighted ES from the fixed-effect model is −0.068 and is statistically very significant. Although this result clearly indicates that the influence of SRI is negative to portfolio performance, it can be interpreted as having a very small effect economically

Table 3 Descriptive statistics of effect sizes

This table shows distributive characteristics of ESs calculated from 205 comparisons. The figures in parentheses are *p*-values. *** indicates statistical significance at the 1% level.

N	Min.	Median	Max.	Mean	Std	Skewness	Kurtosis
205	−1.516	0.015	2.449	0.016 (0.611)	0.455	1.004 (0.000)***	12.755 (0.000)***

Table 4 Weighted effect size

This table shows weighted ESs and its 95% confidence interval from fixed- and random-effect models. It also reports test statistics for the appropriateness of using the random-effect model, Cochran's Q or I^2 statistic. *** represents statistical significance levels at the 1% levels, respectively.

Fixed-effect model			Random-effect model		
Weighted ES	Confidence interval		Weighted ES	Confidence interval	
−0.068***	−0.075	−0.061	0.011	−0.031	0.054
Heterogeneity chi-squared = 4058.62*** (d.f. = 204)			Estimate of between-study variance τ^2 = 0.0764		
p = 0.000					
I^2 (variation in ES attributable to heterogeneity) = 95.0%					
Test of ES = 0 : z = 18.16 p = 0.000			Test of ES = 0 : z = 0.51 p = 0.607		

according to Table 1. Besides, the test statistic to judge between the fixed- and random-effect models, Cochran's Q, strongly suggests that the random-effect model should be used in order to control heterogeneity among different studies. The value of I^2 indicates that 95% of total variance is caused by between-study heterogeneity.

For the random-effect model, the variance used to calculate weight is the sum of within-study variance and between-study variance. In Table 4, the between-study variance is reported as τ^2, 0.0764, which is quite large. Unlike the result of the fixed-effect model, the weighted ES from the random-effect model is positive in sign, but the size is negligible and lacks statistical significance. This result largely supports *Hypothesis 1* and is consistent with the evidence in many empirical papers on SRI (Sauer, 1997; Bello, 2005; Kreander *et al.*, 2005; Bauer *et al.*, 2007; Gregory and Whittaker, 2007).

Although the overall ES is not significantly different from zero, ESs can be systematically different depending on the moderators. We first report the results for moderators associated with sampling issues and proceed to present the results for moderators related to methodological issues. Predictions for moderators and empirical results are summarized in Table 2.

5.2. Results on Sampling Issues

5.2.1. Time.

Table 5 shows AESs and ESs in each decade from the 1970s. ESs are very positive in the beginning. The SRI movement has been initiated by social activists and religious groups who emphasize various merits of SRI. However, the initial positive atmosphere has weakened as more firms engage in CSR activities and SRI performance undergoes more rigorous scrutiny by many researchers and investors, leading to a smaller SRI impact over time. In addition, the impact of SRI tends to decrease

Table 5 Changes in the absolute value of effect sizes

This table reports descriptive statistics for the changes in the absolute value of ESs and ESs themselves through decades. The first column indicates each decade from the 1970s. There is no result for the 1980s because our sample does not have any paper published in the 1980s.

Years	Variable	Obs	Mean	Std dev.	Min.	Max.
1970s	AES	2	0.231	0.050	0.196	0.266
1990s	AES	27	0.324	0.444	0.003	1.516
2000s	AES	119	0.279	0.426	0.001	2.449
2010s	AES	57	0.145	0.200	0.000	1.295
1970s	ES	2	0.231	0.050	0.196	0.266
1990s	ES	27	−0.161	0.528	−1.516	0.617
2000s	ES	119	0.086	0.503	−1.452	2.449
2010s	ES	57	−0.053	0.242	−1.295	0.471

with alternating signs, and both AESs and standard deviation tend to decrease over time. This result supports *Hypothesis 2-1* in that researchers find a smaller impact for SRI over time and the disagreement among them has become smaller, too. The trend of finding a positive impact of SRI is followed by an average negative impact of SRI in the next decade, which is again followed by no significant influence of SRI as expected by dialectical movements of theory development.

The meta-regression results in Tables 6 and 7 further reinforce this interpretation on research trends in SRI area. The negative coefficients for the publication year variable suggest that researchers have come to obtain smaller AESs over time.

Another interesting result regarding the time dimension is shown in Table 8. During an economic crisis, investment performance is lower than it is in a normal period. SRI is not an exception in this regard. The negative coefficients of the Crisis variable indicate that papers whose sample periods include an economic crisis tend to have smaller ESs, which is consistent with *Hypothesis 2-2*.

5.2.2. Publication status.

We run a random-effect meta-regression with AES as the dependent variable and dummies for publication status and statistical significance as independent variables to test *Hypothesis 3*. Univariate (Multivariate) results are shown in Table 6 (Table 7), where tau^2 indicates the restricted maximum likelihood (REML) estimate of between-study variance and I^2_{res} is a percentage of residual variation due to heterogeneity. Since more than 85% of residual variance is caused by heterogeneity, running a random-effect meta-regression is well justified.

Contrary to the prediction of *Hypothesis 3*, the sign for the publication dummy is negative and statistically significant. When combined with the statistical significance dummy, the cross-multiplication term has a negative sign, as shown in model (5) of Table 7. It suggests that among statistically significant comparisons,

Table 6 Impact of sampling moderators on absolute value of effect sizes (univariate)

This table reports univariate meta-regression results about moderators associated with sampling issues. The dependent variable is the absolute value of ESs, and the independent variables are defined in Table 2. The τ^2 indicates the restricted maximum likelihood (REML) estimate of between-study variance, and I^2_{res} means a percentage of residual variation due to heterogeneity. Numbers in parentheses are standard errors. *** and ** indicate statistical significance at the 1% and 5% levels, respectively.

AES	(1)	(2)	(3)	(4)	(5)	(6)
Publication	−0.229 (0.077)***					
Pub_year		−0.008 (0.004)**				
Academia			−0.113 (0.050)**			
Quality				−0.217 (0.047)**		
SRP_CI					0.183 (0.052)***	
SRI_CI						−0.353 (0.169)**
_cons	0.418 (0.074)***	15.917 (7.905)**	0.304 (0.047)***	0.362 (0.040)***	0.165 (0.025)***	0.453 (0.076)***
	τ^2 = 0.069; I^2_{res} = 86.57%	τ^2 = 0.068; I^2_{res} = 86.59%	τ^2 = 0.069; I^2_{res} = 86.57%	τ^2 = 0.062; I^2_{res} = 86.50%	τ^2 = 0.064; I^2_{res} = 86.31%	τ^2 = 0.197; I^2_{res} = 93.44%

Table 7 Impact of sampling moderators on absolute value of effect sizes (multivariate)

This table reports multivariate meta-regression results about moderators associated with sampling issues. The dependent variable is the absolute value of ESs, and the independent variables are defined in Table 2. Numbers in parentheses are standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

AES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Publication	−0.169 (0.072)**				0.009 (0.100)	−0.142 (0.082)*		0.030 (0.209)	
Pub_year	−0.007 (0.004)**	−0.006 (0.004)		−0.007 (0.004)*	−0.007 (0.004)**	−0.003 (0.004)	−0.005 (0.004)	−0.004 (0.009)	
Academia		−0.096 (0.051)*		−0.096 (0.046)**	−0.034 (0.053)		0.022 (0.061)	−0.124 (0.128)	−0.046 (0.047)
Quality			−0.171 (0.045)***						−0.155 (0.049)***
SRP_CI			0.086 (0.049)*			0.134 (0.058)**	0.090 (0.050)*		0.084 (0.049)*
SRI_CI								−0.348 (0.176)*	
Significance	0.251 (0.041)***		0.231 (0.039)***	0.257 (0.041)***	0.554 (0.135)***		0.440 (0.086)***		0.234 (0.040)***
Publication* Significance					−0.332 (0.142)**				
Academia* Significance							−0.230 (0.089)***		
_cons	15.070 (7.231)**	12.769 (8.025)	0.212 (0.045)***	14.550 (7.316)**	14.777 (7.239)*	7.293 (8.118)	10.703 (7.489)	9.226 (18.201)	0.239 (0.052)***
	$\tau^2 = 0.053$; $I^2_{res} = 85.32\%$	$\tau^2 = 0.067$; $I^2_{res} = 86.69\%$	$\tau^2 = 0.047$; $I^2_{res} = 84.96\%$	$\tau^2 = 0.052$; $I^2_{res} = 85.29\%$	$\tau^2 = 0.052$; $I^2_{res} = 85.36\%$	$\tau^2 = 0.065$; $I^2_{res} = 86.47\%$	$\tau^2 = 0.049$; $I^2_{res} = 84.90\%$	$\tau^2 = 0.204$; $I^2_{res} = 93.56\%$	$\tau^2 = 0.048$; $I^2_{res} = 85.00\%$

Table 8 Impact of sampling moderators on effect sizes

This table reports meta-regression results for moderators associated with sampling issues. The dependent variable is ESs, and the independent variables are defined in Table 2. Numbers in parentheses are standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

ES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SRP_CI	−0.113 (0.066)*					−0.144 (0.071)**	−0.160 (0.068)**	−0.152 (0.075)**	−0.165 (0.075)**	−0.174 (0.074)**
Domestic		0.247 (0.128)*					0.478 (0.147)***	0.393 (0.151)***	0.428 (0.151)***	0.445 (0.149)***
Screening			0.114 (0.050)			0.083 (0.050)*	0.067 (0.050)	0.075 (0.051)	0.061 (0.051)	0.064 (0.051)
Mutual				−0.154 (0.056)***		−0.148 (0.061)**	−0.154 (0.059)**		−0.123 (0.063)*	−0.140 (0.061)**
Crisis					−0.274 (0.126)**			−0.215 (0.126)*	−0.151 (0.129)	−0.140 (0.129)
Quality						−0.095 (0.069)		−0.037 (0.073)	−0.044 (0.072)	−0.040 (0.072)
Fin_Jnl						0.061 (0.062)		0.094 (0.060)	0.061 (0.062)	
_cons	0.043 (0.032)	−0.230 (0.131)*	0.002 (0.029)	0.084 (0.037)**	0.030 (0.029)	−0.133 (0.084)**	−0.372 (0.147)**	−0.372 (0.147)**	−0.329 (0.172)*	−0.304 (0.170)*
	$\tau^2 = 0.120$; $I^2_{res} = 91.1\%$	$\tau^2 = 0.113$; $I^2_{res} = 91.2\%$	$\tau^2 = 0.120$; $I^2_{res} = 91.3\%$	$\tau^2 = 0.115$; $I^2_{res} = 90.0\%$	$\tau^2 = 0.114$; $I^2_{res} = 89.2\%$	$\tau^2 = 0.117$; $I^2_{res} = 90.3\%$	$\tau^2 = 0.106$; $I^2_{res} = 90.4\%$	$\tau^2 = 0.107$; $I^2_{res} = 89.6\%$	$\tau^2 = 0.106$; $I^2_{res} = 89.0\%$	$\tau^2 = 0.106$; $I^2_{res} = 89.0\%$

published papers have a smaller impact on AESs. This is again inconsistent with the argument of publication bias.

In addition, we perform a test for publication bias using the Egger test of Stata 14.0 and report the result in Table 9. Since the intercept is not significantly different from zero, we cannot reject the null hypothesis of no publication bias. Therefore, we can conclude that publication bias is not a serious concern for our paper.

The negative coefficient for the publication variable may indicate the quality of papers. Since authors have to go through a rigorous reviewing process to publish their papers in good journals, overly affirmative or overly negative results will be toned down. Thus, if the reviewing process is more rigorous, the AES will become smaller. To examine this idea, we introduce a new variable, Quality, which captures the quality of papers. It is equal to one for journals above median ranking in the same discipline classified by Journal Citation Reports of Clarivate Analytics, and zero for journals below median ranking. As can be seen in Tables 6 and 7, the coefficients of the Quality variable are negative, indicating that papers published in better quality journals tend to have smaller AESs.

5.2.3. Author type.

Academic researchers and industry professionals are different in many respects, including incentives for research (Vaihekoski, 2004). In academic circles, the efficient market hypothesis is widely accepted, but industry professionals are more concerned with finding superior abnormal return opportunities than providing evidence that supports the efficient market hypothesis. This difference will have a dampening effect on AESs for academic researchers. Consistent with this argument, Tables 6 and 7 show that academic researchers tend to find significantly smaller AESs than industry professionals, supporting *Hypothesis 4-1*. When combined with the statistical significance dummy in model (7) of Table 7, the cross-multiplication term of the two variables shows a significantly negative sign. It implies that among significant results, researchers in academia tend to find lower AESs than industry professionals, which is again consistent with *Hypothesis 4-1*.

It is possible that researchers in different disciplines may place a different priority on social and economic values. Researchers in the finance area may put a higher priority on profit maximization than social aspects, producing results showing a

Table 9 Publication bias

This table shows the result of the publication bias test using Stata 14.0. The dependent variable is a standard normal deviate of intervention effect, and the independent variable is a measure of precision, the reciprocal of standard error. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Std_Eff	Coef.	Std err.	t	$p > t$	[95% conf. interval]	
Slope	−0.085***	0.020	−4.21	0.000	−0.125	−0.045
Bias	0.574	0.379	1.51	0.132	−0.173	1.320

lower performance of SRI. On the other hand, business ethics scholars may think social benefit is far more important than profit maximization and generate results showing a higher SRI performance. However, the results in Table 8 are not consistent with this argument and *Hypothesis 4-2* is not supported.

5.2.4. Control group.

In general, SRI funds characteristics are not the same as those of the market index, which includes not only SRI funds stocks but also stocks of all other conventional funds. In addition, since SRI cannot have unconstrained access to whole investment opportunities, SRI portfolios are different from well-diversified market portfolios. Owing to this difference, *Hypothesis 5-1* predicts that papers comparing SRI portfolios with the market index will produce larger AESs than papers comparing SRI and conventional portfolios composed by author(s) according to their research objectives. Consistent with this hypothesis, the dummy variable for market index as a control group shows a significantly positive sign in Tables 6 and 7. This argument is further reinforced by the negative sign of the SRI_CI variable, which is consistent with *Hypothesis 5-2*. Since the difference is lower when comparing two indexes than when comparing SRI portfolios with the index, the former tends to have smaller AESs.

On the other hand, Table 8 shows meta-regression results on the influence of sampling moderators upon ESs themselves. Consistent with *Hypothesis 5-3*, the dummy variable for using the market index as a control group shows a significantly negative sign. It implies that the relative loss of diversification is larger when SRI portfolios are compared with the market index rather than other control portfolios. The negative sign is also consistent with Luther and Matatko (1994) who obtain mostly negative excess returns of SRI portfolios when they use the market index as benchmark.

5.2.5. Investment universe.

In Table 8, the dummy variable for the domestic market shows a positively significant coefficient, which is consistent with *Hypothesis 6*. Since the cultural aspect is an important determinant of ESs, the effect of SRI will be stronger when the investment universe is domestic rather than international. Markets with a homogeneous cultural background will be affected more strongly by SRI, since fund managers will adopt investment strategies that incorporate values appreciated in their culture (Louche and Lydenberg, 2006).

5.2.6. Screening procedure.

Table 8 shows that SRI funds using a positive screening procedure outperform funds with a negative screening, as postulated by *Hypothesis 7*. This result is consistent with Humphrey and Lee (2011) who find that a positive screening significantly reduces funds' risk, but a negative screening increases risk since it impairs the ability of funds to form diversified portfolios. More directly, Goldreyer and Diltz (1999) and Statman and Glushkov (2009) show that a positive screening has a positive impact on portfolio performance, but a negative screening has a negative impact.

5.2.7. *Mutual funds.*

Performance of mutual funds is subject to various factors such as transaction costs, fees and expenditures, and managerial compensation that are largely ignored in analyses on portfolios composed by researchers. In addition, it is well known that managed mutual funds do not perform better than the market as a whole (Jensen, 1968; Carhart, 1997). Therefore, ESs on SRI mutual funds will be smaller than ESs on SRI portfolios formed by researchers. A significantly negative sign for the coefficient of mutual fund moderator in Table 8 is supportive of these arguments and is consistent with *Hypothesis 8*.

5.3. Results on Methodological Issues

In this section, we report the impact that empirical methodologies used by author (s) may have on the AES. Table 10 (Table 11) contains the results of univariate (multivariate) regressions.

5.3.1. *Variable adjustments.*

To begin with, the coefficient of the risk adjustment dummy is statistically significant with a negative sign. This is consistent with *Hypothesis 9-1*, indicating that the AESs calculated from risk-adjusted performance measures are systematically smaller than those calculated from raw returns. Table 12 reports the risk-adjusted performance measures used in the sample. The majority of papers use Jensen’s alpha, and only a small number of papers use the Sharpe ratio or the Treynor index (M’Zali and Turcotte, 1998). Other measures are mostly based on accounting data such as return on assets and the average ES is very large (Cohen *et al.*, 1997). Basically, stock returns are forward looking, but accounting data are not, and it is not uncommon to have conflicting results in finance research. Moreover, the small sample size of other measures makes it difficult draw any serious inference.

Since SRI funds have a different exposure from conventional funds to small firms, equally weighted portfolio returns can generate undeserved performance differences between the two, unlike value-weighted returns. In Tables 10 and 11, the coefficient of weighting scheme has a negative sign and is statistically significant, indicating that the value weighting scheme tends to have smaller AESs, as predicted by *Hypothesis 9-2*.

Hypothesis 9-3 suggests that using more accurate data can reduce the possibility of obtaining a spurious difference between SRI and conventional investments. For example, if SRI funds invest more in small firms than conventional funds, ignoring dividends may produce unwarranted differences in SRI and conventional portfolio performance. Therefore, papers improving data accuracy by considering dividends and stock splits when calculating returns will have smaller AESs. The significantly negative sign of data accuracy variable in model (2) of Table 11 is consistent with *Hypothesis 9-3* although other regressions do not show statistical significance.

Table 10 The impact of methodological issues (univariate)

This table reports univariate meta-regression results about the impact of methodological refinement on SRI research. The dependent variable is an absolute value of ESs, and the independent variables are defined in Table 2. The numbers in parentheses are standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

AES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Risk_Adj	−0.092 (0.047)*						
Index		−0.076 (0.046)*					
Weighting			−0.115 (0.051)**				
Matching				−0.095 (0.054)*			
Match_N					−0.025 (0.016)		
Conditional						0.103 (0.142)	
Accuracy							0.055 (0.053)
_cons	0.267 (0.040)***	0.205 (0.034)***	0.248 (0.028)***	0.246 (0.029)***	0.238 (0.056)***	0.208 (0.029)***	0.194 (0.026)***
	$\tau^2 = 0.060$; $I^2_{\text{res}} = 85.02\%$	$\tau^2 = 0.039$; $I^2_{\text{res}} = 69.85\%$	$\tau^2 = 0.070$; $I^2_{\text{res}} = 85.88\%$	$\tau^2 = 0.075$; $I^2_{\text{res}} = 87.11\%$	$\tau^2 = 0.011$; $I^2_{\text{res}} = 52.24\%$	$\tau^2 = 0.077$; $I^2_{\text{res}} = 87.89\%$	$\tau^2 = 0.067$; $I^2_{\text{res}} = 86.96\%$

Table 11 The impact of methodological issues (multivariate)

This table shows multivariate meta-regression results for the impact of methodological refinement on AESs. The dependent variable is an absolute value of ESs, and the independent variables are defined in Table 2. The numbers in parentheses are standard errors. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

AES	(1)	(2)	(3)	(4)
Risk_Adj	−0.166 (0.091)*	−0.178 (0.074)**		
Index			−0.172 (0.069)**	−0.029 (0.066)
Weighting	−0.146 (0.075)**	−0.207 (0.089)**	−0.110 (0.078)	−0.165 (0.086)*
Matching	−0.159 (0.069)**		−0.155 (0.076)**	
Match_N		−0.032 (0.018)*		−0.015 (0.017)
Conditional	0.075 (0.147)	0.008 (0.131)	0.143 (0.147)	0.044 (0.116)
Accuracy	0.004 (0.076)	−0.194 (0.070)***	0.003 (0.084)	−0.096 (0.084)
_cons	0.461 (0.092)***	0.454 (0.077)***	0.384 (0.079)***	0.217 (0.078)**
	$\tau^2 = 0.077$; $I^2_{res} = 83.11\%$	$\tau^2 = 0.001$; $I^2_{res} = 31.80\%$	$\tau^2 = 0.067$; $I^2_{res} = 76.70\%$	$\tau^2 = 0.001$; $I^2_{res} = 12.66\%$

Table 12 Performance measures used in sampled papers

This table shows descriptive statistics on performance measures used in the sample papers. There are three risk-adjusted performance measures derived from return data and other measures derived from mostly accounting data. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	Obs	Mean	Std dev.	Min.	Max.
Jensen's alpha	126	0.007	0.277	−1.452	0.630
Sharpe ratio	11	−0.066	0.331	−0.581	0.390
Treynor index	5	−0.387	0.621	−1.161	0.075
Others	13	0.906***	0.885	−0.100	2.449

5.3.2. Refining benchmark model.

Some authors use a single-index model as a return-generating process to compute excess returns (Hamilton *et al.*, 1993; Reyes and Grieb, 1998; Goldreyer and Diltz, 1999; Statman, 2000; Bello, 2005), while others use multi-factor models (Luther *et al.*, 1992; Luther and Matatko, 1994; Gregory *et al.*, 1997). The negative sign on the Index variable in Tables 10 and 11 suggests that the AESs calculated using multi-factor models are smaller than those calculated using a single-index model, consistent with *Hypothesis 10.* Since multi-index models control for more dimensions of corporate characteristics than single-index models, a spurious difference in performances between SRI and conventional portfolios tends to be smaller than in the case of using a single-index model.

In practice, fund managers change their portfolio holdings if macroeconomic conditions change, resulting in dynamic risk–return profiles. In this case, conditional models are more appropriate for handling dynamic return-generating processes (Christopherson *et al.*, 1998; Chong *et al.*, 2006), and studies using conditional models are expected to have lower AESs. However, the result lacks a statistical significance, not supporting *Hypothesis 10*.

5.3.3. Matching.

Matching procedures will eliminate the spurious difference between SRI and conventional performance caused by confounding factors and induce AESs to be smaller. Consistent with this argument postulated in *Hypothesis 11*, Tables 10 and 11 show that the coefficient of matching dummy has a significantly negative sign. In addition, the negative sign of the Match_N variable suggests that possible bias caused by the peculiarity of a matching firm in one-to-one matching can be minimized by matching multiple counterparts, leading to a reduction in unwarranted differences between SRI and conventional portfolios.

6. Conclusions and Discussions

This article investigated the impact of SRI on investment performance in the US, using a meta-analysis. The meta-analysis makes a comparison across diverse studies possible by computing ES that is a standardized measure of impact derived from treatment and control groups.

The empirical results show that the weighted average ES is not different from zero. It is an important finding for investors since they do not have to be concerned with the possibility of sacrificing returns when they pursue investment strategies that are socially responsible.

We establish various moderators and run meta-regressions in order to examine whether ESs and AESs change systematically depending on research characteristics. For this, we split moderators into two groups, one related to sampling issues and the other related to methodological issues. To begin with, we find that ESs are positive in the earlier period and have become smaller in terms of absolute value with alternating signs as time passes. When the idea of SRI was relatively new, investors seemed to be able to earn excess returns by incorporating SRI elements in their portfolio. However, the excess return opportunity may have decreased over time and researchers seem to find a smaller impact of SRI since spurious differences between SRI and conventional investment performances are eliminated by the improvement in sampling and methodology and this pattern is by and large consistent with a dialectical development of theories. We also find that papers with sample periods containing economic crisis tend to show a lower ES, as expected.

As for publication status, we find no evidence of publication bias, so we can interpret our results without this complication caused by the publication status of papers. An interesting finding is that better quality journals show relatively smaller AESs than lower quality journals, suggesting that rigorous review processes make papers tone down unwarranted excessive claims.

Whereas the efficient market hypothesis is widely accepted in academia, researchers from industry are more interested in developing a logical backup for industry practices such as active portfolio management and finding the profit potential in SRI. Owing to this difference in research incentives and environment, academic researchers tend to uncover smaller AESs than industry professionals. However, there is no systematic difference in SRI performance among scholars in different disciplines.

Obviously, the chance of having spurious performance difference between SRI and conventional funds becomes higher as the two groups are more different. Consistent with this conjecture, we find that the AESs are bigger when SRI samples are compared with market indexes than when SRI samples are compared with conventional portfolios. And, the AESs are smaller when SRI indexes, not SRI portfolios, are compared with conventional indexes.

Studies comparing SRI portfolios with market portfolios rather than other conventional portfolios tend to show lower ESs. In addition, studies examining negative screening tend to generate smaller ESs than those with positive screening. SRI portfolios and negative screening portfolios seem to sacrifice a diversification benefit, compared with market portfolio and positive screening portfolios, respectively.

Cultural background seems to be an important factor, since domestic SRI funds tend to show higher ESs than international funds. Finally, studies of mutual fund samples show smaller ESs than those of author-composed portfolios. This is largely consistent with capital market efficiency and evidence on mutual fund performances that managed funds rarely beat the market. This result is also related to the fact that mutual fund performance incorporates all costs such as fees, transaction costs, and management compensation, whereas a comparison of composed portfolios in general does not fully take this aspect into account.

We investigate three methodological issues: variable adjustment, the refining benchmark model, and matching procedure. Regarding variable adjustment, our results indicate that any spurious difference in performance between SRI and conventional portfolios can be reduced by using risk-adjusted performance measures, value-weighting schemes, and more accurate variables. Refining benchmark models is related to the issue of return-generating models. Papers using multi-factor models seem to report a lower SRI impact than those using a single-index model. As for matching procedure, papers employing a matching strategy between SRI and control portfolios tend to show smaller AESs. These results suggest that researchers must try to secure similarities as much as possible to minimize a spurious performance difference between SRI and control portfolios. Researchers also have to avoid making various methodological errors.

There is a limitation to drawing a universal conclusion on the impact of SRI upon investment performance, since our empirical result is based only on the US market. However, the findings in this article can be a valid starting point for future research. A natural candidate for future research is an inter-country or interregional comparison of SRI performances. Although data collection and refinement and the interpretation of results based on different cultural backgrounds would be quite challenging, such a study will definitely enhance our understanding of this important issue.

As for methodology, it is true that meta-analysis is a good tool to summarize many papers on the same subject. However, a wide variety of differences among selected papers in terms of sampling and methodology is a significant challenge when performing a sound meta-analysis. Theoretically, the SRI performance can be different depending on the research period, crisis versus normal times, security types, portfolio diversification, social and cultural backgrounds, domestic versus international investments, and so forth. However, even though there would be only one true level of performance difference between SRI and control portfolios caused by a particular factor, researchers may find a dissimilar performance difference if they use different research methodologies. Therefore, future research should be able to separate out more clearly a true SRI impact and the variation in observed SRI performance caused by differences in the research methodologies employed.

References

Adamsson, H., and A. G. F. Hoepner, 2015, *The 'price of sin' aversion: Ivory tower illusion or real investable alpha?* SSRN. Available from: https://ssrn.com/abstract=2659098

Alexander, G. J., and R. A. Buchholz, 1978, Corporate social responsibility and stock market performance, *Academy of Management Journal* 21, pp. 479–486.

Auer, B., 2016, Do socially responsible investment policies add or destroy European stock portfolio value? *Journal of Business Ethics* 135, pp. 381–397.

Ayadi, M., H. Ben-Ameur, and L. Kryzanowski, (2011). Luck versus skill in the cross-section of ethical mutual funds (Les Cahiers du GERAD, Canada).

Barnett, M. L., and R. M. Salomon, 2006, Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance, *Strategic Management Journal* 27, pp. 1101–1122.

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

Bauer, R., J. Derwall, and R. Otten, 2007, The ethical mutual fund performance debate: New evidence from Canada, *Journal of Business Ethics* 70, pp. 111–124.

Becchetti, L., and R. Ciciretti, 2009, Corporate social responsibility and stock market performance, *Applied Financial Economics* 19, pp. 1283–1293.

Bello, Z. Y., 2005, Socially responsible investing and portfolio diversification, *Journal of Financial Research* 28, pp. 41–57.

Benson, K., and J. E. Humphrey, 2008, Socially responsible investment funds: Investor reaction to current and past returns, *Journal of Banking and Finance* 32, pp. 1850–1859.

Benson, K., T. Brailsford, and J. Humphrey, 2006, Do socially responsible fund managers really invest differently? *Journal of Business Ethics* 65, pp. 337–357.

Bollen, N. P. B., 2007, Mutual fund attributes and investor behavior, *Journal of Financial and Quantitative Analysis* 42, pp. 683–708.
Borgers, A., J. Derwall, K. Koedijk, and J. Horst, 2015, Do social factors influence investment behavior and performance? Evidence from mutual fund holdings, *Journal of Banking and Finance* 60, pp. 112–126.
Brander, J., 2006, The effect of ethical fund portfolio inclusion on executive compensation, *Journal of Business Ethics* 69, pp. 317–329.
Capelle-Blancard, G., and S. Monjon, 2014, The performance of socially responsible funds: Does the screening process matter? *European Financial Management* 20, pp. 494–520.
Carhart, M. M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, pp. 57–82.
Chang, C. E., and H. D. Witte, 2010, Performance evaluation of U.S. socially responsible mutual funds: Revisiting doing good and doing well, *American Journal of Business* 25, pp. 9–21.
Chegut, A., H. Schenk, and B. Scholtens, 2011, Assessing SRI fund performance research: Best practices in empirical analysis, *Sustainable Development* 19, pp. 77–94.
Chong, J., M. Her, and G. M. Phillips, 2006, To sin or not to sin? Now that's the question, *Journal of Asset Management* 6, pp. 406–417.
Christopherson, J. A., W. E. Ferson, and D. A. Glassman, 1998, Conditioning manager alphas on economic information: Another look at the persistence of performance, *Review of Financial Studies* 11, pp. 111–142.
Climent, F., and P. Soriano, 2011, Green and good? The investment performance of US environmental mutual funds, *Journal of Business Ethics* 103, pp. 275–287.
Cohen, J., 1998, Statistical power analysis for the behavioral sciences, Routledge. ISBN 978-1-134-74270-7.
Cohen, M. A., S. A. Fenn, and S. Konar, 1997. Environmental and financial performance: Are they related? Working Paper, Vanderbilt University, Nashville, TN.
Combs, J. G., D. J. Ketchen Jr, T. R. Crook, and P. L. Roth, 2011, Assessing cumulative evidence within 'macro' research: Why meta-analysis should be preferred over vote counting, *Journal of Management Studies* 48, pp. 178–197.
Core, J. E., W. R. Guay, and T. O. Rusticus, 2006, Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations, *Journal of Finance* 61, pp. 655–687.
Derwall, J., and K. Koedijk, 2009, Socially responsible fixed-income funds, *Journal of Business Finance and Accounting* 36, pp. 210–229.
Derwall, J., N. Guënster, R. Bauer, and K. Koedijk, 2004, The eco-efficiency premium puzzle, Working Paper, Rotterdam School of Management, Erasmus University, Netherlands.
Derwall, J., N. Guënster, R. Bauer, and K. Koedijk, 2005, The eco-efficiency premium puzzle, *Financial Analysts Journal* 61, pp. 51–63.
Di Bartolomeo, D., and L. Kurtz, 1999, Managing risk exposures of socially screened portfolios, Working Paper, Northfield Information Services, Boston.
Diltz, J. D., 1995, The private cost of socially responsible investing, *Applied Financial Economics* 5, pp. 69–77.
Durand, R. B., S. Koh, and M. Limkriangkrai, 2013, Saints versus sinners. Does morality matter? *Journal of International Financial Markets, Institutions and Money* 24, pp. 166–183.

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

Fabozzi, F. J., K. C. Ma, and B. J. Oliphant, 2008, Sin stock returns, *Journal of Portfolio Management* 35, pp. 82–94.

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

Ferruz, L., F. Muñoz, and M. Vargas, 2012, Managerial abilities: Evidence from religious mutual fund managers, *Journal of Business Ethics* 105, pp. 503–517.

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

Ferson, W. E., and V. A. Warther, 1996, Evaluating fund performance in a dynamic market, *Financial Analysts Journal* 52, pp. 20–28.

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

Geczy, C., R. F. Stambaugh, and D. Levin, 2005, Investing in socially responsible mutual funds, Working Paper, Wharton School, University of Pennsylvania.

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

Girard, E., H. Rahman, and B. Stone, 2007, Socially responsible investments: Goody-two-shoes or bad to the bone? *Journal of Investing* 16, pp. 96–110.

Goldreyer, E. F., and J. D. Diltz, 1999, The performance of socially responsible mutual funds: Incorporating sociopolitical information in portfolio selection, *Managerial Finance* 25, pp. 23–36.

Gompers, P., J. Ishii, and A. Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics* 118, pp. 107–155.

Gregory, A., and J. Whittaker, 2007, Performance and performance persistence of 'ethical' unit trusts in the UK, *Journal of Business Finance and Accounting* 34, pp. 1327–1344.

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

Guenster, N., J. Derwall, R. Bauer, and K. Koedijk, 2005, The economic value of corporate eco-efficiency, Working Paper, Erasmus University.

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

Hasselblad, V., and L. V. Hedges, 1995, Meta-analysis of screening and diagnostic tests, *Psychological Bulletin* 117, pp. 167–178.

Havemann, R., and P. Webster, 1999, Does ethical investment pay? (Ethical Investment Research Service (EIRIS), London). Available from: http://www.eiris.org/files/research%20publications/doesethicalinvestmentpay99.pdf

Hedges, L. V., 1981, Distribution theory for Glass's estimator of effect size and related estimators, *Journal of Educational Statistics* 6, pp. 107–128.

Hickman, K. A., and W. R. Teets, 1999, Social investing and modern portfolio theory, *American Business Review* 17, pp. 72–78.

Hill, R., T. Ainscough, T. Shank, and D. Manullang, 2007, Corporate social responsibility and socially responsible investing: A global perspective (01674544). Available from: http://

offcampus.lib.washington.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=23591308&site=ehost-live

Hoepner, A., and L. Schopohl, 2018, On the price of morals in markets: An empirical study of the Swedish AP-Funds and the Norwegian Government Pension Fund, *Journal of Business Ethics* 151, pp. 1–28.

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

Humphrey, J. E., and D. D. Lee, 2011, Australian socially responsible funds: Performance, risk and screening intensity, *Journal of Business Ethics* 102, pp. 519–535.

Humphrey, J. E., and D. Tan, 2014, Does it really hurt to be responsible? *Journal of Business Ethics* 122, pp. 375–386.

Humphrey, J. E., G. J. Warren, and J. Boon, 2016, What is different about socially responsible funds? A holdings-based analysis, *Journal of Business Ethics* 138, pp. 263–277.

Hutton, R. B., L. D'Antonio, and T. Johnsen, 1998, Socially responsible investing, *Business and Society* 37, pp. 281–305.

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

Jones, S., S. van der Laan, G. Frost, and J. Loftus, 2008, The investment performance of socially responsible investment funds in Australia, *Journal of Business Ethics* 80, pp. 181–203.

Kacperczyk, M., C. Sialm, and Z. Lu, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance* 60, pp. 1983–2011.

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

Kempf, A., and P. Osthoff, 2008, SRI funds: Nomen est omen, *Journal of Business Finance and Accounting* 35, pp. 1276–1294.

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

Lipsey, M. W., 2001, Practical meta-analysis (Sage Publications, Thousand Oaks, CA).

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

Louche, C., and S. Lydenberg, 2006, Socially responsible investment: Differences between Europe and United States, Working Paper, Vlerick Leuven Gent Management School.

Luther, R. G., and J. Matatko, 1994, The performance of ethical unit trusts: Choosing an appropriate benchmark, *British Accounting Review* 26, pp. 77–89.

Luther, R. G., J. Matatko, and D. C. Corner, 1992, The investment performance of UK "ethical" unit trusts, *Accounting, Auditing and Accountability Journal* 5, pp. 57–70.

Margolis, J., H. Elfenbein, and J. Walsh, 2009, Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. http://dx.doi.org/10.2139/ssrn.1866371

Moskowitz, M. R., 1972, Choosing socially responsible stocks, *Business and Society Review*, 1, 71–75

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

Muñoz, F., M. Vargas, and I. Marco, 2014, Environmental mutual funds: Financial performance and managerial abilities, *Journal of Business Ethics* 124, pp. 551–569.

Muñoz, F., R. Vicente, and L. Ferruz, 2015, Stock-picking and style-timing abilities: A comparative analysis of conventional and socially responsible mutual funds in the US market, *Quantitative Finance* 15, pp. 345–358.

M'Zali, B., and M. F. Turcotte, 1998, The financial performance of Canadian and American environmental and social mutual funds, paper presented at the 7th International Conference on The Greening of Industry Network, Rome, Italy.

O'Connor, K., 2003, Dialectic (The Chicago School of Media Theory, Keywords Glossary).

Peiris, D., and J. Evans, 2010, The relationship between environmental social governance factors and U.S. stock performance, *Journal of Investing* 19, pp. 104–112.

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, pp. 349–363.

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

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

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

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

Sauer, D. A., 1997, The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini Equity Mutual Fund, *Review of Financial Economics* 6, pp. 137–149.

Sawilowsky, S., 2009, New effect size rules of thumb, *Journal of Modern Applied Statistical Methods* 8, pp. 467–474.

Schröder, M., 2003, Socially responsible investments in Germany, Switzerland and the United States: An analysis of investment funds and indices, ZEW Discussion Paper.

Shank, T. M., D. K. Manullang, and R. P. Hill, 2005, Is it better to be naughty or nice? *Journal of Investing* 14, pp. 82–87.

Smith, M. P., 1996, Shareholder activism by institutional investors: Evidence from CalPERS, *Journal of Finance* 51, pp. 227–252.

Song, F., A. J. Eastwood, S. Gilbody, L. Duley, and A. J. Sutton, 2000, Publication and related biases: A review, *Health Technology Assessment* 4, pp. 1–115.

Stanley, T. D., 2005, Beyond publication bias, *Journal of Economic Surveys* 19, pp. 309–345.

Statman, M., 2000, Socially responsible mutual funds, *Financial Analysts Journal* 56, pp. 30–39.

Statman, M., 2006, Socially responsible indexes, *Journal of Portfolio Management* 32, pp. 100–109.

Statman, M., and D. Glushkov, 2009, The wages of social responsibility, *Financial Analysts Journal* 65, pp. 33–46.

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

Vaihekoski, M., 2004, Portfolio construction for tests of asset pricing models, *Financial Markets, Institutions and Instruments* 13, pp. 1–39.

Wu, M. L., 2006, Corporate social performance, corporate financial performance, and firm size: A meta-analysis, *Journal of American Academy of Business* 8, pp. 163–171.

Zivney, T. I., and W. J. Bertin, 1992, Publish or perish: What the competition is really doing, *Journal of Finance* 47, pp. 295–329.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. Supplementary information.

Figure S1. Flow chart for paper selection process.