Meta-Analysis on SRI

Chang-Soo Kim,
Faculty of Business Administration,
Yonsei University, South Korea.
E-mail: kimc@yonsei.ac.kr

Abstract
This paper performs a meta-analysis on the impact of SRI upon investment performance based on 205 US samples. The result shows that weighted average effect size is not significantly different from zero, suggesting that investment performance of SRI is not different from conventional investments. Meta-regression results show that, regarding sampling issues, studies published, written by scholars, with domestic investment universe, and with positive screening have higher effect sizes, while studies written by influential authors, comparing with market index, and examining mutual fund samples show lower effect sizes. As for methodology, studies with conditional benchmark and dividend inclusion have larger effect sizes.

Key Words: Effect size, meta-analysis, meta-regression, random effect model, socially responsible investment

JEL Classification: G11, G32, M14
1. Introduction

Socially responsible investment (SRI) has long been a favorite subject since it has opened a possibility of earning money while doing good. Early studies date back to 1970s (Moskowitz 1972, Vance 1975, Alexander and Buchholz 1978), although major research endeavor has started from 1990s. By far, a significant amount of research has been accumulated with a wide range of subjects among which the most interesting and intriguing one is the performance of SRI over conventional investments. Researchers have examined if SRI outperforms conventional investments, since if it is so it will have a profound implication for investment practices and capital market policies.

Existing literature is mixed with positive, negative, and no effect views all together as for the impact of SRI on investment performance. For example, Luther et al. (1992) find weak evidence that SRI funds outperform conventional funds. Mallin et al. (1995) also find evidence that ethical trusts outperform non-ethical trusts in the UK even though they improve on methodological issues to overcome performance measurement problem in the previous literature. However, Gregory et al. (1997) find that the outperformance of UK ethical funds disappears when two-factor benchmark is used to control size effect. Scholtens (2005) uncovers that Dutch SRI funds outperform conventional funds, but the difference lacks statistical significance. Besides, there are many other papers that report no difference in performance between SRI and conventional funds (Hamilton et al. 1993; Sauer 1997; Statman 2000; Bauer et al. 2005, 2006; Bello 2005; Kreander et al. 2005; Scholtens 2005; Benson et al. 2006; Gregory and Whittaker 2007; Kempf and Osthoff 2008).

As an example of underperformance of SRI, Renneboog et al. (2008b) report that SRI funds of France, Ireland, Sweden, and Japan significantly underperform conventional funds by 4–7% per annum during the period of 1991–2003. Stenström and Thorell (2010) also report inferior performance of SRI funds to that of regular funds.

However, there has been almost no attempt to synthesize existing literature on SRI except a limited number of papers. For example, Rathner (2003) attempts to summarize existing literature, but it lacks representativeness due to a small sample size and is incomplete in terms of methodology. Another reason of not having sound summary papers so far is a diversity and complexity of existing papers with regards to samples, methodologies employed, performance measures, investment universe, and benchmarks, to list just a few.

A meta-analysis is a very useful scientific tool to synthesize a wide variety of heterogeneous papers. It calculates effect size (ES) that can be interpreted as a standardized difference between performances of SRI and conventional funds. By standardizing performance differences whatever the performance measures are, it generates a common
measure of difference between the two types of funds making synthesis possible across very diverse types of papers.

In this paper, we carefully explain a meta-analysis and data collection process to secure a replicability of the study. The result on weighted average ES indicates that SRI consideration in investment strategy does not have a significant impact on investment performance in either positive or negative direction. Even so, the results from meta-regression based on various moderators suggest a variety of interesting implications for both academic research and practice.

This study is organized as follows: In the next section, we examine theoretical background regarding the impact of SRI on portfolio performance. This will help us to put in perspective what we find in the empirical section. Section 3 explains a meta-analysis methodology and the concept of ES in detail. We also present a sampling process step by step, since sampling is crucial in performing and interpreting a meta-analysis. Section 4 reports the meta-analysis results on overall value of ESs and the influence of various moderators on ESs. Finally, conclusions and discussions on our findings and future research agenda are presented in Section 5.

2. Literature Review

There are basically three different arguments regarding the effect of SRI on investment performance, i.e., incorporating social elements into investment strategy can have positive, negative, and no impact on portfolio performance.

2.1 Positive View

An obvious rationale for outperformance comes from the results on corporate social responsibility (CSR) literature which shows that CSR firms in general perform better than non-CSR firms (Wu 2006; Margolis et al. 2007). Since SRI invests more on CSR firms, the performance of SRI will be higher than conventional investments. Similarly, Porter and Kramer (2006) assert that having a high ethical standard can be a source of competitive advantage in the situation where customers value ethical principles. Renneboog et al. (2008b) argue that a better SR performance may signal a better managerial quality which leads to a higher profit of the firm. SRI can also be less prone to volatility in the market since social investors tend to invest on firms that focus more on long-term performance.

The argument for a positive impact of SRI also predicts that the performance of SRI funds will improve over time even though their performance could be lower when SRI concept is not well known to investors. In fact, Bauer et al. (2005, 2006) uncover that the performance of SRI funds improve over time, which can be interpreted as a learning effect. Similarly, Cummings (2000) and Barnett and Salomon (2006) find that the performance of SRI funds is better in the long run than in the short run.
2.2 Negative View

The argument for a lower SRI performance largely hinges on portfolio theory (Markowitz 1952). Other things being equal, putting a SRI constraint on investment universe will result in a less favorable efficient frontier, sacrificing diversification benefit (Girard et al. 2007 and Renneboog et al. 2008b). Clow (1999) asserts that constraints placed on investment opportunity set induce sector bias that will increase risk level and SRI is not an exception. An interesting example of worsening investment opportunity set is so called ‘sin stocks’. Hong and Kacperczyk (2009) contend that sin stocks generate a higher risk adjusted return than conventional stocks due to neglect. Fabozzi et al. (2008) and Statman and Glushkov (2009) also uncover that sin stocks perform better than other stocks. If this is the case, excluding sin stocks will cost investors in terms of returns in exchange for their pursuit of personal values.

An argument based on transaction costs also predicts a negative performance of SRI. The subject of SRI is quite qualitative, so the reporting requirement is not as strict as economic dimensions. This implies that there will be a higher cost in gathering and interpreting information. As for the management of SRI funds, they are relatively small, so there is a loss of economies of scale, which leads to a higher transaction costs and management fees (Bauer et al. 2005 and Barnett and Salomon 2006). Ayadi (2011) finds that the average expense ratio of Canadian SRI fund is 1.97% which is much larger than 1.29% of global benchmark of conventional funds.

2.3 No Effect View

In addition to arguments for a positive and a negative impact of SRI, there are rationales for no difference between SRI and conventional investment performance. A powerful argument for no difference is the well-known efficient market hypothesis. If markets are semi-strong form efficient, any investment strategy based on public information cannot generate abnormal returns. If SRI portfolios and funds produce abnormal returns, investors and fund managers can easily replicate their ethical screening procedures which is public information. Even though some SRI funds outperform other funds, their performance superiority may vanish if transaction costs are subtracted.

The contention based on efficient market hypothesis also predicts that a temporary excess return on SRI funds will eventually disappear. Even though investors may systematically undervalue social responsibility dimensions in the short run, the errors in expectation cannot last long under efficient market. Indeed, Derwall et al. (2011) find that the profit opportunities of SRI funds disappear in the long run.

3. Methodology

3.1 Meta-analysis
Since existing studies show quite conflicting results as for the impact of SRI, providing a bird’s eye view on this subject will help us understand the effect of SRI. Traditionally, either one of the following two methods, i.e., narrative synthesis and vote counting, has been used to synthesize research results. When the narrative synthesis is employed, a knowledgeable researcher in the area proposes a summary paper based on his reading lists. Although useful this method cannot avoid an inherent bias of the researcher, since he/she often has a prior in the subject. Under voting method, existing papers are categorized into two groups based on the statistical significance of a difference between treatment and control groups and synthesized according to the results of a majority group. This method shows information on statistical significance of existing literature, but cannot reflect the magnitude or degree of the difference (Jin, 2014)

A meta-analysis is a scientific tool to synthesis existing literature while getting over weaknesses of the traditional methods. Since a meta-analysis collects data as much as possible to secure representativeness and performs a statistical test, it is not subject to a bias caused by an author’s prior in the narrative method. It also reflects the magnitude of differences between treatment and control groups by calculating effect sizes of individual studies and reporting overall effect size utilizing information on the importance of individual studies. Specifically, the ES of individual experiment is calculated by Cohen’s (1988) $d$ which is similar to Hedges’ (1981) $g$ as follows:

$$g = \frac{\bar{p}_T - \bar{p}_C}{\sigma}$$  \hspace{1cm} (1)

where $\bar{p}_T$ ($\bar{p}_C$) is the average performance of SRI (conventional) group and $\sigma$ is the pooled within-group standard deviation which can be computed using the following formula:

$$\sigma = \sqrt{\frac{(N_T-1)V_T + (N_C-1)V_C}{N_T + N_C - 2}}$$  \hspace{1cm} (2)

where $N_T$ ($N_C$) is the sample size of SRI (conventional) group and $V_T$ ($V_C$) is the variance of SRI (conventional) group. The meta-analysis is especially useful when synthesizing diverse studies in terms of samples, methodologies, and research periods, since the ES is a standardized difference between two groups whatever the performance measures are.

There are many studies that use matched pair comparison in SRI literature. This is quite natural since SRI and conventional portfolios have to be compared over the same sample period. For a matched pair comparison, we calculate ESs in two ways depending on data availability of each study. If correlation coefficients between treatment group and control group in addition to the number of pairs, means, and standard deviations are all provided, the standard deviation used in Equation (1) is calculated as follows:

$$\sigma = \sqrt{V_T + V_C - 2 * \rho * \sqrt{V_T * V_C}}$$  \hspace{1cm} (3)
where $p$ is a correlation coefficient between treatment and control groups. If raw data of performance are provided, the ES is calculated using the following equation.

$$g = \frac{\overline{D}}{\sigma}$$

(4)

where $\overline{D}$ is the average performance difference between SRI and conventional groups and $\sigma$ is the standard deviation of differences.

Hedges’ $g$ obtained above is often biased due to a small sample size. If the sample size is not big enough, we adjust it using the following formulas to obtain an unbiased estimator for independent and paired samples, respectively, as suggested by Hedges & Olkin (1985).

$$d = \left(1 + \frac{3}{4(N^2 + N^C - 1)}\right)g$$

(5)

$$d = \left(1 + \frac{3}{4(N-1}-1\right)g$$

(6)

where $N$ is the number of pairs.

In some papers portfolio performances of SRI and control funds are not given, and instead t-statistic of mean difference is reported. In these cases, if two independent samples are compared, the ES can be derived from t-values using the following equation (Rosenthal, 1984):

$$d = \frac{2\tau}{\sqrt{N^2 + N^C - 2}}$$

(7)

However, if the two samples are correlated as in the case of a paired comparison, the ES is calculated as follows:

$$d = \frac{\tau}{\sqrt{N-1}}$$

(8)

After calculating the ESs of individual comparisons, we compute the weighted ES of the total sample in order to see the overall SRI impact on portfolio performance as proposed by Hasselblad and Hedges (1995). The basic idea of weighting is to give higher weights to studies with more precise ES estimates. For this, weights are obtained by computing the reciprocal of variance for each comparison.

When aggregating individual ESs, two models are considered, i.e., fixed effect model and random effect model. The fixed effect model is used when the ESs from individual studies are relatively homogeneous and the variability among ESs is mostly caused by sampling error (Lipsey and Wilson, 2001). Specifically, the weighted average ES from the fixed-effect model is calculated as follows:

$$\overline{d} = \frac{\sum_{i=1}^{k} d_i w_i}{\sum_{i=1}^{k} w_i}$$

(9)

where $d_i$ is ES from individual comparison $i$, and $w_i$ is a weight on comparison $i$ that is a reciprocal of variance $V_i$. The variance of comparison $i$ is calculated using the following formula.
After computing the weighted average ES of fixed effect model, we test whether the homogeneity assumption is satisfied or not. An intuitive way of checking heterogeneity among individual studies is to examine a forest map that shows confidence intervals of included studies holistically. For a more rigorous test of homogeneity, we use Cochran’s Q test or $I^2$ test that has become more popular recently. $I^2$ indicates variation in ES attributable to heterogeneity and is calculated as ratio of between-study variance to total variance. Differently from Q statistic, it is not affected by the number of studies included in meta-analysis. The value of $I^2$ is interpreted as low heterogeneity (25%), medium heterogeneity (50%) and high heterogeneity (75%) (Higgins et al. 2003). The statistics for Q and $I^2$ are calculated as follows.

$$Q_T = \sum_{i=1}^{k} w_i (ES_i - \overline{ES})^2$$  \hspace{1cm} (11)

$$I^2 = 100 \frac{Q_T - (k-1)}{Q_T}$$  \hspace{1cm} (12)

If the test results reject homogeneity, the random effect model will be used. The weighted average ES from the random effect model is calculated as follows.

$$\overline{d^*} = \frac{\sum_{i=1}^{k} d_i w_i^*}{\sum_{i=1}^{k} w_i^*}$$  \hspace{1cm} (13)

where $d_i$ is ES from individual research $i$, and the weight $w_i^*$ is calculated as follows.

$$w_i^* = V_i + V_i^*$$  \hspace{1cm} (14)

Here $V_i^*$ is calculated using the following formula.

$$V_i^* = \frac{Q_T - k - 1}{\sum_{i=1}^{k} w_i - \left( \frac{\sum_{i=1}^{k} w_i^*}{\sum_{i=1}^{k} w_i} \right)^2}$$  \hspace{1cm} (15)

As a rule of thumb, if the sample size is big enough, analysis based on random effect model is more appropriate. However, if the sample size is small, it is recommended to consider both random and fixed effect models and check whether there is a significant difference in the results.

After investigating the overall ES of SRI, we delve into the possibility that SRI impacts on financial performance may vary systematically according to moderators. For this we perform meta-regression analysis as developed by Lipsey and Wilson (2001). The moderators are generated by considering theoretical determinants and methodological characteristics as explained in the next section.

3.2 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 like EBSCO, JSTOR, Science Direct, Springer Link, Wiley-Blackwell, and SSRN, and utilize a web search engine like Google Scholar. We also attempt to collect papers from the following...

A detailed description of our data collection process is showed in Table 2. We first compose keywords that are related to SRI as extensively as possible. When doing this we take into account a different usage of terms in different regions. For example, in the UK, Australia and Canada, socially responsible funds are designated as ‘ethical funds’, whereas similar funds are in general called as ‘socially responsible funds’ in the US and the rest of Europe (Cortez 2009).

By using a phrase ‘socially responsible investment’, we obtain 1,555 articles that include any word in the phrase. Then we go through all articles one by one to sort out those that have a potential to be included in our final list of papers. By using the next phrase ‘socially responsible investing’ we end up having 1,226 results. Since the two phrases are quite similar, most observations are duplicates of the first search and we obtain only 2 additional articles that have a potential to be included in the final sample. The process goes on with new phrases until we use ‘ethical investing’ as shown in Table 2. Since quite a lot of redundancies occur by that time, we begin to use exact phrase from ‘social investing’ to narrow down the search results. Another important data collection process is to scrutinize the reference section of selected papers, especially summary or synthesis papers (Renneboog *et al.* 2008a; Chegut *et al.*. 2011; Capelle-Blancard & Monjon 2014).

**Table 2: Keyword Search**

This table shows a keyword search process. Column 1 shows keywords and column 2 is the number of search results that include the keywords. The number of search results includes duplicates from previous searches. Column 3 shows the number of additional observations from the next keyword search, column 4 is a database access date, and column 5 shows whether exact phrase is used or not.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>N</th>
<th>dN</th>
<th>Access Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>socially responsible investment</td>
<td>1,555</td>
<td></td>
<td>2016-10-03</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>socially responsible investing</td>
<td>1,226</td>
<td>2</td>
<td>2016-10-03</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>SRI</td>
<td>328</td>
<td>1</td>
<td>2016-10-05</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>RI</td>
<td>85</td>
<td>0</td>
<td>2016-10-05</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>responsible investment</td>
<td>2,315</td>
<td>4</td>
<td>2016-10-05</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>responsible investing</td>
<td>1,794</td>
<td>0</td>
<td>2016-10-06</td>
<td>Inclusive of any word</td>
</tr>
<tr>
<td>Ethical investment</td>
<td>2,856</td>
<td>3</td>
<td>2016-10-06</td>
<td>Inclusive of any word</td>
</tr>
</tbody>
</table>
After potential papers are collected a more rigorous reading is performed to decide the papers that are included in the final sample. Since the quality of data is of utmost importance for an accurate assessment of SRI effects we pay a very careful attention to data compilation process. The following explains a part of steps we have taken to preserve data integrity.

1) Many papers have just discussions without any quantitative data, so we drop all these qualitative analyses.

2) We exclude results with obvious errors. For example, Climent (2011) mentions in Table VII that the values in brackets are t-statistics derived from Newey–West heteroskedasticity and autocorrelation consistent standard errors. However, the values cannot be t-statistics since all figures are less than 0.1, but are reported as being statistically significant. When reporting performance of zero-investment portfolio in their Table 3 Trinks and Scholtens (2015) make a mistake, too. In a subpanel entitled ‘Neg.screened S&P versus Unscreened S&P’ regression coefficients and standard errors are all identical, which is very unlikely to correct results.

3) If there are several versions of same paper, we use the most recent paper to avoid overrepresentation bias. For example, Derwall (2004) is a working paper version of Derwall (2005), and we use the latter.

4) We include studies with matching firms or studies without matching firms but with market indices as benchmarks of SRI performance.

5) We include all relevant analyses based on important dimensions such as investment universe, sub-periods, conditional vs. conditional regression, matching vs. non-matching, value weighted vs equal weighted, risk adjusted vs. risk unadjusted, one factor model vs. multi-factor model, etc. The number of observations for meta-analysis is determined by
the total number of analyses, since each paper can report multiple analyses.

6) Although a paper provides necessary data to calculate ESs including numbers of SRI and conventional funds, these numbers are double-checked to be used to calculate ESs. Sometimes these numbers are inappropriate to calculate ESs given the nature of research design of the paper. For example, Renneboog et al. (2008b) report all data required to calculate ESs, i.e., number of observations, mean, and standard deviation. However, we use t-statistics from a pairwise comparison of means between SRI group conventional group instead of using these numbers, since they compare not cross-sectional averages, but time series averages of the two groups. So, when extracting data necessary to calculate ESs, we read empirical methodology section carefully to figure out whether the number of observations is the number of firms or the number of periods.

7) If raw data are reported and the required information for computing ESs is missing, we calculate the required statistics from raw data. For example, Havemann and Webster (1999) does not show statistics required for meta-analysis but it reports time series of values for ethical and conventional indices. We compute means and standard deviations from these raw data. Similarly, in cases where aggregate performance statistics required to compare treatment and control groups are not reported but performance measures of individual funds are reported, we calculate ES by calculating means and standard deviations obtained from individual fund data.

8) If there are results on sub-periods, these are included and the result for the whole period is excluded. This is to exclude a situation where the sub-period results are significant but the whole period result is insignificant. For example, if one sub-period has a negatively significant result and the other sub-period has a positively significant result, then the whole period result tends to be insignificant due to averaging out effect. Besides, inclusion of both whole period and sub-period results is subject to the error of double counting. Similarly, if a paper reports results on multiple countries, only US market results are included.

9) If authors run many similar regressions by using different criteria, only the results that authors indicate to be more relevant or important are included. For example, Statman and Glushkov (2009) show many performance results based on various criteria such as sub-period, social characteristics, weighting schemes, and indices. We include only total period results since authors do not have any specific rationale to separate the whole period into two sub-periods. But we include results based on weighting scheme, i.e., equally weighted and value weighted, since weighting scheme is one of moderator variables for our meta-regression. Similarly, when regression results generated by a various combination of independent variables are reported, only the most representative result is included. For example, Luther and Matatko (1994) present several tables that
show results from running regressions using various combinations of two market indices, but the results are very similar. We include only the two index model which is said to be most appropriate by authors.

10) We pay a special attention to the sign of ESs when coding. For example, Fabozzi et al. (2008), Hong and Kacperczyk (2009), and Adamsson and Hoepner (2015) study the impact of sin stocks by comparing the portfolio of sin stocks with conventional portfolio. To be consistent with other papers that examine good stocks, i.e., socially responsible stocks, we multiply figures of those papers by (-1).

11) When coding dummy variable, we carefully distinguish between ‘none’ and ‘not applicable’. For example, Ferruz (2012) employs conditional model where he uses four conditioning variables in addition to multi-factors. When composing data from his descriptive statistics about raw returns (Table 1), we code the value of dummy variable for conditional model as ‘na’, not ‘0’, since the use of conditional model is only a relevant issue when comparing risk adjusted factor models such as four factor Carhart model, not when comparing raw returns.

After all these screening process, we end up having 205 analyses from 51 papers that study the impact of SRI on a financial performance in the US. Our number of observation is the number of experiments, not papers. For example, if a paper reports two tables of comparison between SRI and conventional funds, one is based on raw return and the other based on Jensen’s alpha, we end up having two observations. Since a paper may report 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.

4. Results and Discussion

4.1 Weighted Average ES

We first report the result on overall ES and proceed to show the influence of various moderators on ESs. Table 3 shows 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 Hamilton et al. (1993), Reyes and Grieb (1998), Goldreyer and Diltz (1999), Statman (2000), and Bello (2005). They find no significant difference in the performance of SR and conventional funds in the US market, even though they examine different samples and time periods. The distribution of ESs is not normally distributed since skewness and kurtosis are significantly different from normal distribution.

Table 3: Descriptive Statistics of Effect Sizes

This table shows distributive characteristics of ESs calculated for 205 comparisons. The figures in parentheses are p-values. *** indicates statistical significance of 1% level.
The first result of interest in our meta-analysis is the overall value of ESs that allows us to judge about an average impact of SRI consideration on investment performance. In order to calculate the weighted ES we use the reciprocal of variance of each comparison calculated by Equation (10). When aggregating ESs, we have to decide which model is more appropriate between fixed effect model and random effect model considering sample characteristics.

Table 4 shows a weighted ES from running a fixed effect model. The weighted ES is -0.068 and is statistically very significant, resulting in 95% confidence interval that does not include 0. Although this result clearly indicates that the influence of SRI is negative to portfolio performance, the weighted ES itself can be interpreted as having a very small effect (Sawilowsky, 2009). Besides, the test statistic to judge between fixed and random effect model, Cochran’s Q, strongly suggests that the random effect model should be used instead of fixed effect model 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.

**Table 4: Weighted Effect Size Using Fixed Effect Model**

This table shows weighted ES and its 95% confidence interval when using fixed effect model. It also reports test statistics for appropriateness of using random effect model, Cochran’s Q or $I^2$ statistic. "***", **", *" represent a statistical significance level of 1%, 5%, and 10%, respectively.

<table>
<thead>
<tr>
<th>Weighted ES</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.068***</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>-0.061</td>
</tr>
</tbody>
</table>

Heterogeneity chi-squared = 4058.62 (d.f. = 204) p = 0.000

$I^2$ (variation in ES attributable to heterogeneity) = 95.0%

Test of ES=0 : z=18.16 p = 0.000

Table 5 shows a weighted ES from running random effect model. For random effect model the variance used to calculate weight is the sum of within-study variance and between study variance as shown in Equations (13) (14) and (15). In Table 5 the between study variance is reported as $r^2$, 0.0764, which is quite large. The weighted ES from the random effect model is positive in sign differently from the result based on fixed effect model. But the size is negligible and lacks statistical significance. The 95% confidence interval includes 0.

**Table 5: Weighted Effect Size Using Random Effect Model**

This table shows weighted effect size and its 95% confidence interval when using random effect model. It also reports test statistics for appropriateness of using random effect model, Cochran’s Q or $I^2$ statistic. "***", **", *" represent a statistical significance level of 1%, 5%, and 10%, respectively.

<table>
<thead>
<tr>
<th>Weighted ES</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.068***</td>
<td>-0.075</td>
</tr>
<tr>
<td></td>
<td>-0.061</td>
</tr>
</tbody>
</table>

Heterogeneity chi-squared = 4058.62 (d.f. = 204) p = 0.000

$I^2$ (variation in ES attributable to heterogeneity) = 95.0%

Test of ES=0 : z=18.16 p = 0.000
Heterogeneity chi-squared = 4058.62 (d.f. = 204) p = 0.000
I² (variation in ES attributable to heterogeneity) = 95.0%
Estimate of between-study variance τ² = 0.0764
Test of ES = 0 : z = 0.51 p = 0.607

This is largely consistent with many empirical papers on SRI. For example, Gregory et al. (1997) find that the outperformance of UK ethical funds found in previous papers disappears when they use a two-factor benchmark to control size effect. Gregory and Whittaker (2007) also find that neither ethical nor conventional funds underperform the market as a whole. Scholtens (2005) examines the performance of Dutch socially responsible funds and reports an insignificant outperformance of SRI funds over conventional funds. Kreander et al. (2005) uncover similar performances of SRI and conventional funds when they investigate SRI funds of seven European countries (Belgium, Germany, Netherlands, Norway, Sweden, Switzerland and UK). Bauer et al. (2006) and Bauer et al. (2007) investigate the samples of Australian and Canadian ethical funds, respectively, and again find no statistical difference between the performance of SRI and conventional funds. Other than the above mentioned papers, there are still many papers that report no difference between SRI and conventional investments (Hamilton et al. 1993; Sauer 1997; Statman 2000; Bello 2005; Benson et al. 2006; Kempf and Osthoff 2008; Renneboog et al. 2008b). On balance, judging from the weighted ES based on random effect model, we can conclude that there is almost no difference between the performance of SRI and non-SRI portfolios.

4.2 Results on Moderators

Although the overall ES is not significantly different from zero, there is still a possibility that ESs are systematically different along with moderators. To be consistent with the moderator section, we first report the results on moderators associated with sampling issues and proceed to present the results on moderators related to methodological issues. We examine each moderator one by one, since they are mostly dummy variables and dumping so many dummy variables in one regression does not make much sense. Besides, meta-regression is different from subgroup analysis since it is basically a random effect model.

4.2.1 Sampling

As explained in methodology section, we split the sampling issues into three categories, sampling issues of our paper and sampled papers, and portfolio characteristics of sampled studies.

4.2.2 Moderators for Current Paper

Since we collect studies both published and non-published, there can be difference in ESs between the two groups of studies. So, we run a random effect meta-regression with ESs as dependent variable and the publication status as independent variable. Table 6 shows that the
The coefficient of publication dummy variable is significantly positive indicating that published papers tend to have larger ESs. Other things being equal, larger ESs are equivalent to higher statistical significance. This result is consistent with Song et al. (2000) who argue that papers with a high statistical significance are more likely to be published than papers with insignificant results. In the table $\tau^2$ indicates the restricted maximum likelihood (REML) estimate of between-study variance and $I^2_{\text{res}}$ means a percentage of residual variation due to heterogeneity. Since about 95 percent of residual variance is caused by heterogeneity, running a random effect meta-regression is very well justified.

**Table 6: Impact of Publication Status on Effect Sizes**

This table shows the meta-regression result to see whether there is a systematic difference in ESs depending on a publication status. The dependent variable is ESs and the independent variable is a dummy variable for publication status defined as follows: Publication = 1 if a paper is published and Publication = 0 if a paper is not published. The $\tau^2$ indicates the restricted maximum likelihood (REML) estimate of between-study variance and $I^2_{\text{res}}$ means a percentage of residual variation due to heterogeneity. ***, **, * indicate a statistical significance of 1%, 5%, and 10% level, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication</td>
<td>0.219**</td>
<td>0.091</td>
<td>2.4</td>
<td>0.017</td>
<td>0.039</td>
</tr>
<tr>
<td>_cons</td>
<td>-0.186**</td>
<td>0.087</td>
<td>-2.14</td>
<td>0.033</td>
<td>-0.368</td>
</tr>
</tbody>
</table>

$\tau^2 = 0.105; I^2_{\text{res}} = 94.99\%$

### 5. Researcher Background

There can be a difference in ESs between academic researchers and industry professionals since they are different in many respects inclusive of incentives of research, data accessibility, and transaction and monitoring costs (Vaihekoski 2004). Panel A of Table 7 shows that academic researchers tend to find significantly higher ESs than industry professionals. In Panel B we report the meta-regression result of using both dummies for publication status and author type. Consistent with results of publication status and author type, these two variables have a positive sign which is statistically very significant. The cross-multiplication term of the two variables shows a significantly negative sign, which implies that among published papers researchers in academia report lower ESs than industry professionals.

**Table 7: Impact of Author Type on Effect Sizes**

This table shows meta-regression results to see whether there is a systematic difference in ESs between authors from academia and those from industries in Panel A. The dependent variable is ESs and the independent variable is a dummy variable for author types defined as follows: Academia = 1 if author(s) is from academia, and 0 if author(s) is from industry. In Panel B, we report the influence of author type on ESs in association with publication status of papers. The $\tau^2$ indicates the restricted maximum likelihood (REML) estimate of between-study variance and $I^2_{\text{res}}$ means a percentage of residual variation due to heterogeneity. ***, **, * indicate a statistical significance of 1%, 5%, and 10% level, respectively.

#### Panel A

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academia</td>
<td>0.144**</td>
<td>0.060</td>
<td>2.40</td>
<td>0.017</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Panel B

| Coef.     | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----------|-----------|-------|-----|----------------------|
| Publication | 0.336***   | 0.128 | 2.62| 0.009 | 0.083 | 0.589 |
| Academia   | 0.322***   | 0.096 | 3.37| 0.001 | 0.134 | 0.511 |
| Publication*Academia | -0.442*** | 0.137 | -3.23| 0.001 | -0.712 | -0.172 |
| _cons     | -0.193     | 0.088 | -2.20| 0.029 | -0.366 | -0.020 |

\tau^2 = 0.112; \text{I}_2\text{res} = 91.01%  

6. Number of Papers

It is conceivable that our results on ESs are affected more significantly by influential researchers. Table 8 reports that the number of papers has a negative coefficient that is statistically significant at 10% level, suggesting that active researchers in SRI area tend to report a less strong results of socially responsible investing. The SRI movement has been initiated by social activists and religious groups who emphasize various merits of SRI without paying much attention to financial aspect. However, the initial affirmative atmosphere has weakened as SRI performance has been under more rigorous scrutiny by many researchers, professionals and investors. The above result is consistent with this movement in SRI industry. In other words, as a researcher studies more about SRI, he/she becomes less affected by arguments that support SRI movement unless backed by evidence.

Table 8: Number of Papers and Effect Sizes

This table reports meta-regression results to see whether there is a relationship between the number of papers written by an author(s) and ESs. The dependent variable is ESs and the independent variable is the number of papers. The \( \tau^2 \) indicates the restricted maximum likelihood (REML) estimate of between-study variance and \( \text{I}_2\text{res} \) means a percentage of residual variation due to heterogeneity. ***, **, * indicate a statistical significance of 1%, 5%, and 10% level, respectively.

| Coef.    | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-----------|-------|-----|----------------------|
| Number   | -0.050*   | 0.029 | -1.74| 0.084 | -0.106 | 0.007 |
| _cons    | 0.100***  | 0.056 | 1.78 | 0.077 | -0.011 | 0.210 |

\tau^2 = 0.118; \text{I}_2\text{res} = 91.00%  

6.1 Moderators Related to Sampled Papers

Since existing studies vary widely in terms of many different dimensions, it is worthwhile to investigate the impact of these dimensions on ESs. We consider ones related to sampling in this subsection.
Survivorship bias is often considered to cause a distortion in portfolio performances (Brown et al. 1992). The relative performance of SRI funds over conventional funds would be negatively affected if dead funds are not included in the sample and the failure rate of conventional funds are higher than SRI funds (Kempf and Osthoff 2008). Table 9 reports the result from meta-regression on survivorship bias. It turns out that the coefficient of survivorship bias dummy variable is not statistically significant at all, suggesting that ESs are not affected in any significant way even though researchers do not collect failing funds in their sample.

### Table 9: Impact of Survivorship Bias on Effect Sizes

This table shows meta-regression results to see if there is a systematic difference in ESs depending on the correction of survivorship bias. The dependent variable is ESs and the independent variable is a dummy variable defined as follows: Survivorship = 1 if a paper includes dead funds in its sample, and 0 if a paper does not include dead funds. The \( \tau^2 \) indicates the restricted maximum likelihood (REML) estimate of between-study variance and \( I^2_{\text{res}} \) means a percentage of residual variation due to heterogeneity. ***, **, * indicate a statistical significance of 1%, 5%, and 10% level, respectively.

<table>
<thead>
<tr>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survivorship</td>
<td>-0.029</td>
<td>0.064</td>
<td>-0.45</td>
<td>0.653</td>
</tr>
<tr>
<td>_cons</td>
<td>0.023</td>
<td>0.037</td>
<td>0.63</td>
<td>0.528</td>
</tr>
</tbody>
</table>

\( \tau^2 = 0.130; I^2_{\text{res}} = 90.81\% \)

Since the ES is calculated by comparing treatment and control groups, there can be a systematic difference in ESs depending on control groups. Table 10 shows meta-regression results on the influence of control sample upon ESs. In Panel A the dummy variable for market index as a control group shows a significantly negative sign, which implies that experiments using market index as a comparison group to SRI funds tend to have lower ESs. This result is consistent with the explanation that, since SRI cannot form a fully diversified portfolio, or cannot have unconstrained access to whole investment opportunities, it is subject to a less favorable risk-return trade-off (Hong and Kacperczyk 2009; and Statman and Glushkov 2009).

In general, characteristics of SRI funds are not the same as those of market portfolio that includes not only stocks of SRI funds but also stocks of all other conventional funds in the market. The negative sign is consistent with Luther et al. (1992) and Luther and Matatko (1994) who uncover that SRI mutual funds invest more on small firms than conventional funds. They obtain mostly negative excess returns when they use broadly diversified market index as benchmark, but the signs of excess returns change to mostly positive when they include small firm index together with market index.
In Panel B, the dummy variable for matching procedures does not show a statistically significant result. Even though control samples are composed more rigorously by matching, it does not seem to exert a systematic influence on ESs.

Table 10: Impact of Control Sample on Effect Sizes

This table reports meta-regression results on the issue of control samples. ESs may change systematically depending on characteristics of control samples. The dependent variable is ESs and in Panel A the independent variable is a dummy variable defined as follows: Mkt_Index = 1 if control sample is conventional market index, and 0 if control sample is portfolios or mutual funds. Panel B reports the impact of matching procedure upon ESs and independent variable is defined as follows: Matching = 1 if SRI portfolios are matched with conventional ones with similar characteristics, and 0 if no matching procedure is employed. The tau² indicates the restricted maximum likelihood (REML) estimate of between-study variance and I²_res means a percentage of residual variation due to heterogeneity. ***, **, * indicate a statistical significance of 1%, 5%, and 10% level, respectively.

| Panel A. Use of Market Index | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-----------------------------|-------|-----------|---|------|------------------|
| Mkt_Index                   | -0.113| 0.066     | -1.72 | 0.086 | -0.243, 0.016    |
| _cons                       | 0.043 | 0.032     | 1.33 | 0.184 | -0.021, 0.107    |

\[ \text{tau}^2 = 0.120; \text{I}^2_{\text{res}} = 91.10\% \]

| Panel B. Use of Matching Procedure | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-----------------------------------|-------|-----------|---|------|------------------|
| Matching                          | -0.049| 0.068     | -0.72 | 0.471 | -0.184, 0.085    |
| _cons                            | 0.028 | 0.036     | 0.78 | 0.435 | -0.043, 0.100    |

\[ \text{tau}^2 = 0.132; \text{I}^2_{\text{res}} = 91.72\% \]

7. Conclusions and Recommendations

This paper investigates the impact of SRI consideration on investment performance in the US using a meta-analysis which is a powerful tool especially when existing results have a considerable degree of heterogeneity in terms of samples, research period, and methodologies. It makes comparison across diverse studies possible by computing ES that is a standardized measure of impact derived from treatment and control group.

The empirical results show that the weighted average ES is not different from zero, implying that investors do not have to be concerned with a possibility of losing returns when they pursue investment strategies based on their personal values. At least, SRI does not penalize investors although they cannot earn better returns over conventional investments.

We establish various moderators and run meta-regressions in order to figure out whether there is any systematic change in ESs depending on moderators. To begin with, we find that published papers have higher ESs than unpublished ones, which confirms a so-called publication bias. Since so many papers are submitted, journals become selective in such a way that they in general accept papers with a higher statistical significance. This practice
could lead to an undesirable situation where good papers die out due to a weak statistical significance. Our result shows that ESs tend to be higher if a paper is published even among the group of papers showing a statistical significance.

Another interesting result is that papers written by researchers in academia have high ESs than ones written by industry professionals. Obviously, scholars and professionals have difference incentives for research, access to dataset, exposure to various organizational constraints, etc. This difference may at least partially have an influence on the result. However, when combined with publication status, papers written by scholars tend to have lower ESs than peers written by professionals among published papers. We also check the chance that our result is driven by influential authors. The number of papers written by an author(s) has a negative relation with ESs, suggesting that influential authors tend to produce results that negatively compromise the SRI impact on investments. This result is consistent with an argument that more knowledgeable researchers are less prone to optimism of SRI advocated by activists and provide a dampened view on the impact of SRI.

When SRI is compared with market indexes that are broadly diversified, the performance of SRI is worse than market index. This result suggests several points of discussion. One is that a proper comparison has not been made since SRI portfolios can be different from market indexes in its portfolio composition. By the way, since SRI is less well diversified than market indexes, it is theoretically obvious that SRI risk-return profiles are inferior to market indexes (Hong and Kacperczyk, 2009; and Statman and Glushkov, 2009). The other is that SRI performance has been heavily influenced by small capitalization firms that underperform market indexes (Luther et al. 1992 and Luther and Matatko 1994). However, we cannot find any difference in ESs when using dummy variables for survivorship bias and matching procedures, although these are the ways to improve matching quality between SRI and conventional groups.

Since SRI is based on value system of a society, the performance of SRI will be stronger when its investment universe is more homogeneous in social and cultural background. Our result that portfolios investing on domestic market outperform international portfolios is consistent with this argument. The result on screening method that positive screening procedure produces better performance than negative screening is another evidence that supports the argument based on portfolio diversification loss. As another sampling issue, if authors examine mutual fund samples, they tend to obtain smaller ESs. This is largely consistent with capital market efficiency and evidence on mutual fund performances.

The selection of benchmark return generating process is crucial, since excess returns have to be obtained first to compare SRI and conventional investment performances. We consider three issues regarding benchmark, i.e., risk adjustment, multi-factor models, and conditional models and obtain a systematic difference in ESs only in the case of using conditional models.
as a benchmark return generating process. This result is consistent with behavior of value pursuing investors who do not respond as frequently as conventional investors to changes in macroeconomic conditions if these changes are just of temporary nature.

As for the refinement of input data, using after-fees data does not produce any significant change in ESs. This result is consistent with many empirical results that fees and expenses are not different between SRI and conventional funds (Statman 2000, Bauer et al. 2005, Benson et al. 2006, Benson and Humphrey 2008, and Renneboog et al. 2008b, and Gil-Bazo et al. 2010). However, inclusion of dividends affects ESs negatively, suggesting that dividend payments are lower for SRI funds than conventional funds. This is consistent with underperformance of smaller firms (Luther et al. 1992 and Luther and Matatko 1994). As for the difference in ESs based on either value-weighting or equal-weighting scheme, we cannot find any significant difference in ESs. This may be due to the fact that same weighting scheme is applied to both SRI and conventional samples, making a net impact negligible.

We have to admit that there is a limitation on drawing a universal conclusion on the impact of SRI upon investment performance, since our empirical result is based on only US market. However, it is also true that there are many interesting findings in this paper and this findings can be a good starting point for future research. A natural candidate for future research is an inter-country or inter-region comparisons of SRI performances. Although data collection and refinement would be quite challenging, such study will definitely be a significant contribution to enhancement of our understanding on this important issue.

References
The papers included in the meta-analysis are marked with *.


---

There are in general multiple analyses in one paper. We use the terms ‘analysis’, ‘experiment’ and ‘comparison’ interchangeably.