

Does it Really Hurt to be Responsible?

Jacquelyn E. Humphrey · David T. Tan

Received: 8 November 2012 / Accepted: 4 May 2013 / Published online: 15 May 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Prior literature on socially responsible investment has contended that excluding “sin stocks” from a portfolio (negative screening) will reduce performance and increase risk. Further, incorporating stocks of firms with positive social responsibility scores (positive screening) will improve performance and reduce risk. We simulate portfolios designed to mimic typical equity mutual funds’ holdings and investigate these propositions. We remove the potentially confounding influences of differences in manager skill, transaction costs and fees, and conduct a clean experiment on the effect of positive and negative portfolio screening. We find no difference in the return or risk of screened and unscreened portfolios. We conclude that a typical socially responsible fund will neither gain nor lose from screening its portfolio.

Keywords Socially responsible investing · Ethical investing · Mutual funds

Introduction

Socially responsible investing (SRI), where non-financial information is incorporated into investment decision making, is by no means a new concept. For centuries, groups of

religious investors have chosen to form their investment portfolios in ways that reflect their values (US SIF 2012). Today, there is a much more heterogeneous group of participants in the SRI market, with SRI now incorporating considerations such as climate change, pollution and executive remuneration. Indeed, SRI can no longer be considered a small, niche market—with \$3.74 trillion, or over 11 % of all assets under management, currently invested in this way (US SIF 2012).

Socially responsible investing works by screening portfolios for particular non-financial issues. Negative screening is where firms engaging in undesirable activities, such as gambling or alcohol production, are excluded from the investment portfolio. Positive screening incorporates firms with desirable characteristics, such as good labour relations and involvement with community development, into the portfolio. In this article, we investigate whether positive and negative screening have differing impacts on investment portfolios’ risks and returns.

The easiest and most convenient way for an individual investor to access SRI is via an SRI fund. The US SIF (2012) reports that in the US, the most common screens used by SRI funds are the traditional “sin” screens: tobacco, alcohol and gambling.¹ That is, most of the money invested in SRI funds in the US is negatively screened. However, the majority of SRI funds also use positive screening.² Indeed, most SRI funds use multiple screening criteria and incorporate both positive and negative screens.

J. E. Humphrey (✉) · D. T. Tan
Research School of Finance, Actuarial Studies and Applied
Statistics, Australian National University, Level 4, College of
Business and Economics Building 26C, Canberra, ACT 0200,
Australia
e-mail: jacquelyn.humphrey@anu.edu.au

D. T. Tan
School of Aviation, The University of New South Wales, Old
Main Building, Second Floor, Sydney, NSW 2052, Australia
e-mail: david.tan@unsw.edu.au

¹ The majority of funds also currently screen for firms which are involved in Sudan (US SIF 2012). We include involvement in Sudan as one criterion in our “human rights” screen.

² See <http://ussif.org/resources/mfpc/screening.cfm> for a list of which screens SRI funds use—date accessed 27 July 2011.

There is now an extensive literature which investigates whether there is a cost or a benefit to investing in SRI funds relative to non-SRI, or “conventional”, funds. The overwhelming evidence suggests that there is no difference in the risk-adjusted returns of SRI funds compared to conventional funds (see, for example, Goldreyer and Diltz 1999; Statman 2000; Bauer et al. 2005; Bello 2005; Benson et al. 2006).

More recently, it has been argued that the different types of screens may affect performance and risk in different ways. First, that positive screening results in an increase in returns (Barnett and Salomon 2006). Typically this branch of the literature has used stakeholder theory to justify its claims. In particular, the argument is that firms which engage in positive activities, such as community involvement and good employee relations, are at a competitive advantage to firms which do not. These firms are able to attract superior management and employees and they have good relationships with the communities within which they operate. This then flows through to higher firm profitability and consequently superior returns (Lado and Wilson 1994; Turban and Greening 1997).

However, these arguments are somewhat questionable. If markets are relatively efficient, we should only observe a one-off price increase (increase in return) when the market learns of a firm implementing a particular value-adding responsible practice. For superior returns to persist, either markets would have to exhibit some level of inefficiency and systematically overestimate (underestimate) the benefit (cost) of SRI (Statman and Glushkov 2009), or the firm would have to be constantly innovating in its responsible practices, with each innovation increasing firm value going forward.

It has also been argued that positive screening will lower idiosyncratic risk because responsible behaviour reduces frictions between responsible firms and the societies in which they operate. Examples of frictions would include product boycotts, employee strikes or law suits—such as the public suing a firm for damaging the environment (Cornell and Shapiro 1987). Firms with positive stakeholder relationships would not face these types of frictions, therefore they would experience less shocks to their cash flows and consequently have relatively lower idiosyncratic risk.

In contrast, it has been argued that negative screening reduces returns and increases risk. Returns are reduced because investors forgo potentially profitable opportunities by excluding stocks from their portfolios on non-financial grounds (Adler and Kritzman 2008; Fabozzi et al. 2008; Statman and Glushkov 2009). Indeed, the first two authors conclude their papers with the following dire warnings to investors:

Trustees or fiduciaries who develop institutional investment policy statements should fully understand the economic consequences of screening out stocks of

companies which produce a product that is inconsistent with their value systems. In addition, they should question if the cost to uphold common social standards is worthwhile. (Fabozzi et al. 2008)

A socially motivated investor can...exclude bad (i.e. “sin”) companies from her portfolio and thereby sacrifice vast sums of wealth through time.... (Adler and Kritzman 2008)

In terms of risk, drawing on modern portfolio theory, it has been hypothesized that SRI managers are unable to form fully diversified portfolios because a proportion of the investable universe is excluded (see Barnett and Salomon 2006). This will then result in SRI funds having higher idiosyncratic risk, which is not compensated for by higher return.

As already noted, most SRI funds use a combination of positive and negative screens. Some researchers suggest that any increase in return from positive screening is offset by a corresponding decrease in return from negative screening, which is why prior research has not found evidence of an overall relation between SRI and performance (Statman and Glushkov 2009).

At present, the SRI mutual fund literature has not provided a clean experiment which explicitly examines the differing impact of positive and negative screening on fund performance and returns. Goldreyer and Diltz (1999) examine the difference in the performance of SRI funds which positively screen versus SRI funds which do not and find evidence of positively screened funds outperforming. They do not examine negative screening, or the impact of screening on risk. A number of studies have examined the impact of the *intensity* of screening: whether the number of screens a SRI fund imposes impacts performance and/or risk (Barnett and Salomon 2006; Renneboog et al. 2008; Jegourel and Maveyraud 2010; Lee et al. 2010; Humphrey and Lee 2011). However, examining the impact of the number of screens a portfolio uses is a different question from examining whether positive and negative screening per se affect fund outcomes. Further, this literature has examined actual mutual fund returns. However, the return and risk of mutual funds is determined by a number of factors: the performance of the underlying stocks, the skill of the fund manager, transaction costs and the fees that the fund charges. This means that using actual mutual fund returns may not be the optimal way of determining whether the SRI characteristic has an impact on risk and returns (Schröder 2007; Statman and Glushkov 2009).

There is also a burgeoning literature that examines SRI by forming portfolios based on SRI characteristics, rather than examining mutual funds (see, for example, Galema et al. 2008; Filbeck et al. 2009; Statman and Glushkov 2009). These studies on the whole examine a large universe

of stocks (for example, the entire KLD database) and determine whether screening impacts returns and risk. This analysis is to some extent unrealistic because it is unlikely that a typical SRI investor will hold a portfolio comprising thousands of stocks. We wish to investigate the effect SRI will have on a portfolio a typical retail investor is likely to hold and this is why we choose to simulate portfolios of similar size to a typical mutual fund and examine the effect of positive and negative screening on these portfolios. Mutual funds comprise a significant segment of the SRI market (US SIF 2012) so it is worthwhile investigating screening within the context of mutual funds.

From the risk perspective, we are particularly interested in the impact of SRI on idiosyncratic risk since the theoretical arguments outlined above apply to idiosyncratic risk. Further, an increase in systematic risk is usually associated with a higher expected return, but the idiosyncratic risk is not usually compensated for with a higher return.

In sum, in this article we investigate the effect positive and negative screening has on the performance and risk of SRI funds relative to unscreened funds, removing the confounding effects of managerial skill, fees and expenses.

Our strategy is to simulate portfolios which mimic, as realistically as possible, the characteristics of mutual funds. Since our portfolios comprise the underlying stocks in a fund, and we do not use the actual fund returns, we do not need to be concerned about differences in fees across funds impacting our results. In addition, we remove fund manager skill by randomly selecting stocks when we form the portfolios. This allows us to disentangle the effect of the screening mechanism from fees and skill. Our approach is similar to that of Adler and Kritzman (2008). However, Adler and Kritzman (2008) assume that SRI funds *randomly* eliminate stocks from the portfolio. This approach is incorrect because screening is by no means random: *specific* industries are excluded from SRI portfolios. Further, these authors do not recognize that SRI funds also positively screen their portfolios. Our article overcomes these shortcomings.

We find that positively or negatively screened portfolios do not provide returns or risks that are any different from those of unscreened portfolios. Investors experience neither harm nor benefit from investing in an SRI fund.

The rest of the article proceeds as follows. The data is described in the second section and the methodology in the third section. Results are presented in fourth section and fifth section concludes the article.

Data

Our aim is to form portfolios that are as similar as possible to the actual portfolios held by SRI funds. To this end, we obtain information on SRI funds from a number of sources.

The first step is to determine the initial investible universe of stocks, so we download a list of the benchmarks SRI funds use from the US SIF's website.³ As our analysis will investigate the characteristics of portfolios of stocks, we only investigate funds which use equity benchmarks. The benchmarks which SRI funds use are displayed in Table 1.

The most common benchmark is the S&P 500 total return index, with approximately one-third of funds benchmarking to this index. The next most common index is the Russell 2000 index, which only 11 % of SRI funds benchmark to. The number of funds using a particular benchmark drops rapidly outside of these two indexes. Funds that select stocks from the S&P 500 index obviously have a smaller universe of stocks from which to choose than funds that use the Russell 2000 index, meaning that any constrained diversification effect of SRI will be more pronounced for funds which benchmark to the S&P 500 index. In addition, prior to 2001, our positive screening data is only available for those companies listed on the S&P 500 index.

We check whether the conventional universe is similar by extracting equity funds' primary prospectus benchmark from Morningstar Direct. About 31 % of these funds benchmark to the S&P 500 total return index with the next most popular indexes being the Russell 1000 growth (11 % of funds) and value (9 % of funds) total return indexes. Consequently, it is most logical for us to use the S&P 500 index as our investible universe. In robustness tests, we expand the investible universe to include the 3,000 largest listed US stocks.

We extract the historical list of S&P 500 index constituents, which incorporates all index additions and deletions, from Compustat. This dataset becomes our initial (unscreened) investible universe. Returns and market capitalisations on all stocks on our list of S&P 500 index constituents are extracted from CRSP.

We next need to screen this initial universe of stocks. We use NAICS and SIC codes to identify the stocks to be negatively screened. Since tobacco, alcohol, gambling and defence/weapons are the most widely used negative screens, we exclude stocks in these industries. Specifically, we classify tobacco stocks as those with NAICS codes 312210, 312221 or 312229 or SIC codes 132, 2100–2199, 5194 or 5993. Alcohol stocks have NAICS 312120, 312130 or 312140, or SIC codes 2080–2085, 518, 5181 or 5182. Gambling stocks have NAICS 713210, 713290 or 721120. Defence/weapons stocks have SIC codes 3760–3769 and in this category we also include firearms, which have SIC codes 3480–3489. Our negatively screened universe excludes stocks in any of these classifications.

³ http://ussif.org/resources/mfpc/index.cfm?mf_type=BL&order— date accessed 8 April 2010.

Table 1 SRI fund benchmarks

Benchmark	Number of funds	Percent of funds
S & P 500 Composite Total Return	32	33.3
Russell 2000	11	11.5
Custom	8	8.3
Russell 3000	5	5.2
Calvert Social	4	4.2
MSCI EAFE Equity	4	4.2
Russell Mid Cap Growth	4	4.2
Russell Mid Cap Value	4	4.2
Russell 1000	3	3.1
Russell 2000 Value	3	3.1
Russell 3000 Growth	3	3.1
FTSE KLD 400 Social	2	2.1
MSCI U.S. Prime Market Growth Index	2	2.1
MSCI US Prime Market Value Index	2	2.1
Russell 2000 Growth	2	2.1
Lipper mixed asset target allocation growth funds average	1	1.0
Russell 1000 Value	1	1.0
Russell 2500 Value	1	1.0
Russell 3000 Utilities	1	1.0
Russell Midcap Index	1	1.0
S & P 400 Midcap Total Return	1	1.0
Total number of funds	96	100

This presents information from the Social Investment Forum on the benchmarks used by SRI funds

Classifying sin stocks in this way will only allow us to identify stocks whose predominant business is in a sin industry. We are surprised to discover that only 15 stocks are identified as sin stocks using this criterion over our entire sample period. Indeed, a maximum of only 10 sin stocks are listed on the S&P 500 at any one point in time. However, firm involvement in a particular sin industry may be more indirect. For example, a subsidiary may be involved in a sin industry, or the firm may sell sin products as part of, but not its main, business. Consequently, we also use the MSCI/KLD STAT database (KLD) to identify sin stocks. In particular, we select any firm that KLD identifies as being involved in the following “controversial businesses”: tobacco, alcohol, firearms, gambling, weapons, military and nuclear. On average, 73 stocks per year are excluded from the portfolios using these criteria.

Stocks for our positively screened portfolios are also identified from KLD. KLD rates firms along a number of non-financial criteria, awarding firms a value of one for each “strength” and each “concern”. These strengths and concerns are then aggregated to give a total number of strengths and a total number of concerns across each of the following categories: community, corporate governance,

diversity, employee relations, environment, human rights and product. These categories are similar to the actual screens used by SRI funds, as identified by the US SIF.⁴ In line with prior literature (for example, Statman and Glushkov 2009; Hong and Kostovetsky 2011), we form an overall score for every category by subtracting the total number of concerns from the total number of strengths across each of these seven categories. We then take the total number of strengths minus the total number of concerns and select all firms which have a positive total score, i.e. we select the firms which have more strengths than concerns. This then becomes the universe from which we can form our positive (equally weighted screens) portfolios. However, each category has a different number of elements to it, and these change over time. For example, in 1995, there are six community strengths and four community concerns, but only three corporate governance strengths and three concerns. We therefore also take each firm’s score across each dimension and divide it by the number of strengths/concerns. For example, in 1995 a firm may have five community strengths, which we would then divide by the total number of community strengths available from KLD, which is six. We do this for each strength and concern. We then add each of these weighted scores to get a total score. Firms which have a positive score are then used as the universe from which to form positive (weighted screens) scores.

We use the S&P 500 total return index as our proxy for the market, and in robustness tests we use the CRSP VW index. Return on these indexes and return on the 1-month risk free rate are from the Fama–French Portfolios and Factors database available through WRDS. Our sample period is January 1996–December 2010.

Methodology

We begin our analysis with an investigation of the underlying universes of stocks. This is essentially a replication of previous studies which have looked at screening but have not taken portfolio size into account (for example, Statman and Glushkov 2009). We form five portfolios by value-weighting *all* of the stocks in the underlying universes. No screen comprises all stocks on the S&P 500 total return index. Our negatively screened universes are the S&P 500 constituents with the sin stocks (identified by SIC/NAICS

⁴ Specifically, US SIF identifies the following potential positive screens: environment (comprising climate/clean technology, pollution/toxics and other), social (comprising community development, diversity/equal employment, human rights and labour relations) and corporate governance (comprising board issues and executive pay).

or by KLD) removed.⁵ Positively screened universes are the S&P 500 stocks with positive overall KLD scores.⁶ Correspondence with MSCI/KLD revealed that the ratings are usually released in December of each year and sent to clients at some time in the first quarter. This means that clients would receive, for example, the 1995 KLD ratings at the beginning of 1996. Consequently, to avoid look-ahead bias, we match the KLD ratings with the subsequent year's constituents of the S&P 500, forming portfolios in the January following KLD's data release. We therefore use KLD ratings for the years 1995–2009 and our sample of returns is from January 1996 to December 2010.⁷

We also investigate differences in the universes by subtracting the unscreened universes results from the screened universe results. Results from these difference calculations will indicate whether screening has an impact on performance.

We use a number of performance measures. Jensen's α is the intercept from the following regression:

$$(R_{p,t} - R_{f,t}) = \alpha_p + \beta_p(R_{m,t} - R_{f,t}) + \varepsilon_{p,t} \quad (1)$$

where $R_{p,t}$, $R_{m,t}$, $R_{f,t}$ are the returns on universe p , the market portfolio and the risk-free asset at time t .

We are mainly interested in the results on our difference universes. A significantly positive (negative) α coefficient on our difference universe indicates that screened universes perform better (worse) than the unscreened universe.

The literature has found that there are differences in how socially responsible and socially irresponsible firms load onto the size, book-to-market and momentum risk factors (Bauer et al. 2005; Galema et al. 2008; Statman and Glushkov 2009). If substantial risk premia are available on these factors, factor loadings will impact the returns which could be expected from portfolios. Since our aim is to isolate the impact of positive and negative screening, we need to make sure that any difference in the risk or returns of our portfolios is attributable to the socially responsible characteristic and not differences in these other portfolio

characteristics. Consequently, we also calculate the four-factor alpha as the intercept from the Carhart (1997) model:

$$(R_{p,t} - R_{f,t}) = \alpha_p + \beta_p(R_{m,t} - R_{f,t}) + s_pSMB_t + h_pHML_t + m_pUMD_t + \varepsilon_{p,t} \quad (2)$$

where SMB , HML and UMD are the return on the mimicking size, book-to-market and momentum factors, respectively.

To ensure continuity across the factors, we form SMB , HML and UMD ourselves using only stocks listed on the S&P 500. Returns are obtained from the CRSP database. Market values are calculated as common shares outstanding multiplied by price from the CRSP database. Market values of firms with multiple share issues are calculated using all share classes, but only the most liquid stock's return is utilised. Book value of equity is from Compustat (item CEQ) for the fiscal year t . Negative book values are deleted. Book-to-market ratios are the book value of equity divided by market value.

We then move on to the primary focus of our article, which is to investigate the impact of screening on portfolios which realistically represent what a fund manager could be expected to hold. Clearly, an active manager will not hold the entire universe of stocks. We therefore need to decide how many stocks to have in each portfolio. Morningstar Direct provides a time series of the number of stocks held by mutual funds. We extract this information for domestic equity funds, deleting index funds. Funds hold 122 stocks on average over our entire sample period. If we restrict the funds to only those which benchmark to the S&P 500 total return index, this number drops to 114. Portfolio sizes appear to increase over our time period, ranging from a minimum of 86 (67) to a maximum of 165 (145) stocks for all equity funds (funds which benchmark to the S&P 500). We therefore decide to form portfolios of 100 stocks. We also form portfolios of differing numbers of stocks to investigate whether portfolio size impacts the results.

To create our unscreened portfolios, 100 stocks are randomly selected (without replacement) from our list of S&P 500 constituents in January from 1996 to 2009. The portfolio is held for 12 months. After 12 months, we reform the portfolio from another random draw of 100 stocks. This gives us a portfolio which assumes a 100 % holdings turnover and zero manager skill. Value-weighted portfolio statistics are then calculated for the portfolio over the 15-year period. The process is repeated 10,000 times.

Our screened portfolios are formed in an identical manner, except that the relevant universes from which stocks are bootstrapped are either positively or negatively screened.

⁵ Note that we need to make an assumption about when a portfolio manager would add or delete a stock from the portfolio when the underlying index changes. We assume that deletions to the index are held in the fund until the end of the month and additions are added at the beginning of the following month. This results in our unscreened universe not necessarily comprising exactly 500 stocks in any given month.

⁶ We are unable to test the effect of specific screens due to insufficient numbers of firms available for each screen. For example, there are less than 30 stocks with positive governance scores for half of the years in our sample (in 1 year only 13 stocks have a positive governance score).

⁷ This sample period also has the advantage of allowing us to avoid using KLD's 2010 ratings. In 2009, MSCI acquired RiskMetrics, which had previously acquired KLD and there are noticeable differences in the "strengths" and "weaknesses" criteria between 2009 and 2010.

We now have time series for 10,000 portfolios each of unscreened, positively and negatively screened stocks. We then calculate the difference between pairs of screened and unscreened portfolios—leading to four sets of 10,000 difference observations. We rank the difference portfolios from largest to smallest. To determine whether there is a significant difference between the screened and unscreened portfolios, we create confidence intervals by examining the 5th and 95th percentile values: these are essentially two-tailed 5 % critical values. If both critical values are positive (negative), this means that we can be 95 % confident that screened portfolios have higher (lower) risk or return than unscreened portfolios. However, if one critical value is positive and the other negative, we cannot conclude that there is a difference between the two samples, as zero is contained in the confidence interval.

We perform this procedure for our four definitions of screened portfolios across a number of performance and risk metrics. To investigate fund performance, we first examine raw returns and the Sharpe ratio (returns in excess of the risk-free rate divided by the standard deviation). We then examine the α from a one-factor Jensen model or a four-factor Carhart alpha (Eqs. (1) and (2) above), which adjusts returns for systematic risk.

We measure total risk as the standard deviation of the portfolios' monthly returns over the sample period. We divide total risk into its systematic and idiosyncratic components. We use the beta coefficients from Eqs. (1) and (2) to measure systematic risk. Idiosyncratic risk is the error term from Eqs. (1) or (2).

Results

Results from our investigation of the underlying universes (rather than the bootstrapped portfolios) are in Tables 2 and 3. Descriptive statistics on the entire universes are presented in Table 2. There are considerably less stocks in our positively screened universes than in our unscreened universe, but negatively screened universes are almost as large as our unscreened universe. Returns across the universes are almost identical. We perform t tests on the difference in the raw returns and, unsurprisingly, find no significant difference in the returns of the screened and unscreened universes.⁸ Stocks in our positively screened universes are substantially larger than those in unscreened and negatively screened universes. This highlights the need to use a performance measurement model which takes firm size into account.

Table 3 shows the output from the one- and four-factor models. None of the alphas on the universes are significant,

⁸ Results not displayed, available upon request.

save the difference between the Neg (SIC) and the unscreened universe. We suspect, however, that the significant t statistic is more an artefact of the miniscule standard error (0.000027 and 0.000028 for the one and four-factor models, respectively) rather than denoting any real evidence of underperformance. These tiny standard errors are attributable to the almost identical return series of the two universes (we note that the R^2 of the Neg (SIC) regressions are 0.997; there is essentially no difference between our Neg (SIC) and unscreened universes because such a small number of stocks have been excluded).

To test total risk, we examine differences in the variance of returns of each of our series. F tests indicate that there is no significant difference in the variances of our screened and unscreened universes. Systematic risk is measured as the beta from our one- and four-factor models. Results, in Table 3, show that positively screened and Neg (SIC) universes have higher systematic risk, as measured by beta, while Neg (KLD) have lower betas than the unscreened universes. However, we are mainly interested in the impact of screening on idiosyncratic risk. We measure idiosyncratic risk as the residual from our one- and four-factor models. We test whether screening impacts idiosyncratic risk by performing t tests for differences in the means of our screened and unscreened universes (results not displayed, available upon request). We do not find evidence of screening impacting idiosyncratic risk.

We now turn to our main tests which utilise bootstrapping to simulate portfolios of fund returns. There are now only 100 stocks in each of our screened and unscreened portfolios. Results are presented in Table 4. Performance is evaluated in four different ways: raw returns, Sharpe ratios, one- and four-factor alphas.

Recall that some authors have suggested that positive screening should *increase* returns because the included firms have positive relationships with their stakeholders, which should then result in a higher stock return. We do not find evidence of outperformance in any of our positively screened portfolios using any performance measure.⁹ This finding is in line with prior literature. Statman and Glushkov (2009) do not find significant out- (or indeed under-) performance in portfolios long in stocks with positive KLD scores and short in negative scores.

⁹ It must be noted that the majority of the alphas on our bootstrapped portfolios are insignificant, so it may not be appropriate to use these alphas to calculate differences between the screened and unscreened portfolios. Consequently, we also examine the alphas and replace any insignificant alpha with zero before taking differences. In each case, the upper and lower critical values become zero because the vast majority of the differences are zero. This result still upholds our findings of no difference between screened and unscreened portfolios' alphas. We thank an anonymous referee for raising this point.

Table 2 Universes: descriptive statistics on universes

	No screen	Pos (ES)	Pos (WS)	Neg (SIC)	Neg (KLD)
Number of stocks					
Average	504.61	168.56	107.00	489.00	424.00
Minimum	499	149	174	501	458
Maximum	210	183	146.94	496.03	439.5722
SD	2.15	9.61	15.66	2.25	8.04
Returns					
Average	0.01	0.01	0.01	0.01	0.01
Minimum	-0.21	-0.21	-0.21	-0.21	-0.21
Maximum	0.19	0.18	0.19	0.19	0.18
SD	0.05	0.05	0.05	0.05	0.05
Size					
Average	18,946,298	2,315,0324	20,198,443	1,907,3978	18,427,120
Minimum	8,958,056	10,947,791	10,101,942	8,988,501	8,493,585
Maximum	2,680,2196	33,400,401	2,910,3669	27,036,203	26,776,238
SD	4,288,878	5,897,101	5,136,953	4,344,089	4,357,051

This presents descriptive statistics on the screened and unscreened universes. No screen represents the full sample of all S&P 500 firms. Pos universes are formed from S&P 500 firms which KLD identifies as having strong performance across community, corporate governance, diversity, employee relations, human rights and product. Pos (ES) weights each dimension equally, whereas with Pos (WS), each dimension is weighted according to the number of criteria it encompasses. Neg (SIC) excludes S&P 500 firms whose NAICS or SIC indicate they are from the tobacco, alcohol, gambling or weapons industries. Neg (KLD) excludes firms which KLD identifies as being involved in tobacco, alcohol, firearms, gambling, weapons, military or nuclear. Returns denotes monthly returns, size is calculated as shares outstanding multiplied by price and is in thousands of dollars. The sample period is January 1996–December 2010

However, recall that we have applied the KLD screening information in January following the year in which it is released—the time when most fund managers would receive this data. KLD scores firms using publically available information. Consequently, it is perhaps unsurprising that a positive return cannot be made using this data, since by the time the KLD ratings are made available, the information is already stale. What would happen, however, if managers could identify firms with good social practices *before* the market? We investigate what would happen if we form our portfolios a year prior to the KLD release date. That is, we allow managers to have information about positive screening prior to KLD's public release. However, this does not alter our results. All performance results are still insignificant. We therefore conclude that positive screening, or at least positive screening using KLD data, does *not* provide the market with value-relevant information; even if the information is anticipated prior to it becoming public.

The traditional argument made regarding negative screening is that this process should decrease returns. There is evidence in the literature which suggests that sin stocks deliver superior performance (Fabozzi et al. 2008; Hong and Kasperczyk 2009) and since sin stocks are excluded

from our negatively screened portfolios, we would expect to find underperformance. Further, as we move from using NAICS/SIC classifications to identify sin stocks to using the KLD classifications, we should see a decreasing return as more stocks are excluded from the portfolios. Turning to our results in Table 4, while we do see the average performances of negatively screened portfolios being uniformly less than those of unscreened portfolios, the differences are insignificant. This is an interesting finding because it suggests that negative screening *does not* have a detrimental effect on performance, even though sin stocks are known to outperform.

How can we align this finding with prior literature? Statman and Glushkov (2009) find significant excess returns to portfolios of responsible minus (KLD rated) sin stocks, and Fabozzi et al. (2008) and Hong and Kasperczyk (2009) find significant premiums available for (pure play) sin stocks. However, it must be noted that the excess returns found in Statman and Glushkov (2009) disappear when portfolios are value-weighted. All of these prior studies have a rather large number of sin stocks in their sample. However, as we have documented, there are at best 10 pure play sin stocks available in the S&P 500 at any point in time. While we do not question the finding that sin stocks outperform, our results suggest that their contribution to the returns of a typical mutual fund's portfolio is so small as to be negligible. Consequently, the returns of SRI funds are in no way compromised by excluding these stocks.

Risk results are presented in Table 5 and are broken into three components: total risk (volatility), systematic risk (one- and four-factor beta) and idiosyncratic risk (one- and four-factor standard error). Positively (negatively) screened portfolios have slightly higher (lower) total, systematic risk and tracking error, but there is no *significant* difference in the risk of screened and unscreened portfolios.

However, what we are most interested in is results from idiosyncratic risk. Recall that some have argued that positive screening should result in lower (idiosyncratic) risk as positive stakeholder relations should reduce frictions between a firm and the society within which it operates. In contrast, negative screening should increase portfolio idiosyncratic risk as managers are subject to a restricted investible universe. We do not find evidence in favour of either of these contentions. Screening portfolios does *not* significantly impact on risk.

Robustness Tests

We perform a battery of robustness tests to ensure that our findings are not driven by our methodological choices. Recall that our initial portfolios were formed in January each year. Correspondence with KLD indicated that their

Table 3 Universes: regression output on universes

	One-factor model			Four-factor model					
	α	MRP	Adj. R^2	α	MRP	SMB	HML	UMD	Adj. R^2
No screen	-0.0001	0.99***	1.00	-0.0002	0.99***	0.00	0.03***	0.01	1.00
	-0.88	181.08		-0.99	205.48	0.20	4.06	1.07	
Pos (ES)	-0.0007	1.02***	0.95	-0.0004	1.03***	-0.07**	-0.08***	-0.03	0.95
	-0.88	71.92		-0.59	73.48	-2.01	-2.74	-0.93	
Pos (ES)—no screen	-0.0006	0.03***	0.01	-0.0002	0.04***	-0.07*	-0.11***	-0.03	0.10
	-0.67	2.08		-0.34	3.13	-1.79	-3.96	-1.31	
Pos (WS)	-0.0006	1.03	0.92	-0.0007	1.04***	0.04	-0.16***	-0.02	0.93
	-0.53	41.95***		-0.75	51.38	0.85	-4.18	-0.51	
Pos (WS)—no screen	-0.0004	0.04*	0.01	-0.0005	0.06***	0.04	-0.19***	-0.03	0.17
	-0.39	1.77		-0.56	3.00	0.77	-5.36	-0.70	
Neg (SIC)	-0.0002	0.99***	1.00	-0.0002	0.99***	0.002	0.03***	0.01	1.00
	-1.22	184.12		-1.31	202.13	0.17	4.00	1.17	
Neg (SIC)—no screen	-0.0001***	0.005***	0.26	-0.0001***	0.005***	0.0004	0.0007	0.0006	0.26
	-2.14	5.15		-2.01	5.34	-0.27	-0.80	0.85	
Neg (KLD)	-0.0003	0.97***	0.99	-0.0002	0.97***	-0.02	0.06***	0.00	0.99
	-0.73	88.63		-0.72	105.39	-0.78	4.10	0.08	
Neg (KLD)—no screen	-0.0001	-0.02*	0.02	-0.0001	-0.02**	-0.02	0.03***	-0.01	0.10
	-0.39	-1.82		-0.21	-1.99	-1.35	2.62	-0.37	

This presents regression results on the screened and unscreened universes. Alpha is the intercept, MRP the coefficient on the market risk premium. SMB, HML and UMD are coefficients on the size, book-to-market and momentum factors, respectively. No screen represents the full sample of all S&P 500 firms. Pos universes are formed from S&P 500 firms which KLD identifies as having strong performance across community, corporate governance, diversity, employee relations, human rights and product. Pos (ES) weights each dimension equally, whereas with Pos (WS), each dimension is weighted according to the number of criteria it encompasses. Neg (SIC) excludes S&P 500 firms whose NAICS or SIC indicate they are from the tobacco, alcohol, gambling or weapons industries. Neg (KLD) excludes firms which KLD identifies as being involved in tobacco, alcohol, firearms, gambling, weapons, military or nuclear. The sample period is January 1996–December 2010. Newey–West HAC t statistics are in parentheses

** Significance at the 5 % level

clients receive KLD rating scores in the first quarter of the following year. To mimic a fund manager who may receive the ratings later in the year, we also form portfolios in April. Results are unchanged from our initial specification.

Our original tests were run using a 1-year holding period; so portfolios were rebalanced every year. We also perform the analysis using a 5-year holding period. We find no difference in any of the return or risk measures of our unscreened and screened portfolios using a 5-year holding period.

It is possible that there is a “learning effect” over the sample period where, initially SRI information was not fully understood by the market but over time the market became better at valuing this information and incorporating it into prices.¹⁰ This would manifest in SRI funds exhibiting higher returns in the early stages of our sample period, but the effect tapering out over time. To address this question, we perform 36-month rolling window regressions and examine the one- and four-factor alphas over time. We do not find evidence of a systematic reduction in the

differential alphas of SRI and conventional funds over our sample period.

We had initially formed portfolios of 100 stocks. Given that mutual funds are of differing sizes, we investigate the effect of altering the number of stocks in our portfolios by increasing the number of stocks in increments of 50 up to 400 stocks. We are only able to perform this robustness test for our negatively screened portfolios, since positively screened universes comprise less than 150 stocks in a number of months.¹¹ Altering portfolio size does not impact return or total and systematic risk results. Screened and unscreened portfolios of up to 350 stocks display insignificantly different amounts of idiosyncratic risk. However, portfolios of 400 stocks screened using KLD criteria have significantly higher idiosyncratic risk than unscreened portfolios. What this implies is that the relative

¹⁰ We thank an anonymous referee for this suggestion.

¹¹ We note that positive screening is more likely to be an overlay on an existing diversified portfolio. Our results show that a purely positively screened portfolio is not undiversified. Therefore, we do not expect that adding a positively screened portfolio to an existing diversified portfolio will result in any significant loss of diversification.

Table 4 Performance of samples

	No screen	Pos (ES)	Pos (WS)	Neg (SIC)	Neg (KLD)
Raw returns					
Average value	0.0068	0.0062	0.0064	0.0067	0.0066
Average of differences		-0.0005	-0.0004	-0.00003	-0.0002
Upper critical value		0.0015	0.0016	0.0023	0.0021
Lower critical value		-0.0025	-0.0023	-0.0023	-0.0025
One-factor alpha					
Average value	0.000002	-0.0007	-0.0006	-0.00004	-0.0001
Average of differences		-0.0007	-0.0006	0.00005	-0.0001
Upper critical value		0.0014	0.0015	0.0023	0.0023
Lower critical value		-0.0027	-0.0026	-0.0025	-0.0025
Four-factor alpha					
Average value	-0.00004	-0.0004	-0.0006	-0.0001	-0.0001
Average of differences		-0.0003	-0.0006	-0.00004	0.00004
Upper critical value		0.0017	0.0014	0.0023	0.0023
Lower critical value		-0.0023	-0.0026	-0.0024	-0.0023
Sharpe ratio					
Average value	0.0859	0.0726	0.0737	0.0850	0.0832
Average of differences		-0.0134	-0.0123	-0.0009	-0.0027
Upper critical value		0.0289	0.0298	0.0482	0.0470
Lower critical value		-0.0555	-0.0536	-0.0506	-0.0526

This shows results from examining the returns of 10,000 bootstraps of portfolios of 100 stocks. Upper and lower critical values are the 95th and 5th percentile values of differences between subtracting unscreened from screened portfolios, and provide confidence intervals for our tests. Portfolios are value weighted. No screen represents portfolios formed from the full sample of all S&P 500 firms. Pos portfolios are formed from S&P 500 firms which KLD identifies as having strong performance across community, corporate governance, diversity, employee relations, human rights and product. Pos (ES) portfolios weight each dimension equally, whereas with Pos (WS) portfolios, each dimension is weighted according to the number of criteria it encompasses. Neg (SIC) are formed by excluding S&P 500 firms whose NAICS or SIC indicate they are from the tobacco, alcohol, gambling or weapons industries. Neg (KLD) portfolios are formed from excluding firms which KLD identifies as being involved in tobacco, alcohol, firearms, gambling, weapons, military or nuclear. The sample period is January 1996–December 2010

** Significance at the 5 % level

advantage of an unscreened portfolio only becomes important in a very large portfolio. Perhaps this is not surprising. While in theory, investors should hold all available assets in order to fully diversify away idiosyncratic risk (Markowitz 1952), Statman (2004) asserts that more than 300 stocks are required for mean–variance optimization. Consequently, an unscreened portfolio of 400 stocks is unlikely to have much, if any, residual idiosyncratic risk. Screening this portfolio for sin stocks may therefore result in significantly increased risk. Stated another way, idiosyncratic risk in our context arises from two sources: small portfolio size and screening. In smaller portfolios, both screened and unscreened portfolios will be carrying some diversifiable risk. It seems that the increased idiosyncratic risk from the relatively narrower investible universe for the screened portfolios may be insubstantial relative to the overall increase in systematic risk from having a smaller portfolio. We note, however, that an

actively managed mutual fund with 400 stocks in its portfolio is not usual (recall that that the average number of stocks held by funds which benchmark to the S&P 500 total return index is 114). Stocks using NAICS/SIC criteria have the same amount of idiosyncratic risk as unscreened portfolios even in portfolios of 400 stocks.

It is likely that firms in our screened portfolios may come from different industries from those in our unscreened portfolios. As different industries may have divergent return and risk profiles, it is important to ensure that we are not simply picking up industry characteristics in our screened and unscreened portfolios. We therefore investigate whether industry-adjusting returns impacts on our results. Similar to Edmans (2011), we obtain each stock's SIC industry classification from COMPUSTAT and match these to the 49 Fama and French (1997) industry factors. We then take each firm's return in excess of its industry return and regress these against MRP, SMB, HML

Table 5 Risk of samples

	No screen	Pos (ES)	Pos (WS)	Neg (SIC)	Neg (KLD)
Volatility					
Average value	0.0485	0.0501	0.0513	0.0487	0.0479
Average of differences		0.0016	0.0028	0.0002	-0.0007
Upper critical value		0.0051	0.0064	0.0042	0.0032
Lower critical value		-0.0020	-0.0008	-0.0038	-0.0046
Tracking error					
Average value	0.0485	0.0501	0.0513	0.0487	0.0479
Average of differences		0.0008	0.0029	-0.00004	0.0004
Upper critical value		0.0035	0.0059	0.0030	0.0036
Lower critical value		-0.0019	-0.0001	-0.0030	-0.0027
One-factor beta					
Average value	0.9858	1.0163	1.0301	0.9904	0.9690
Average of differences		0.0305	0.0444	0.0046	-0.0168
Upper critical value		0.1048	0.1177	0.0913	0.0688
Lower critical value		-0.0450	-0.0292	-0.0813	-0.1016
Four-factor beta					
Average value	0.9817	1.0252	1.0389	0.9865	0.9635
Average of differences		0.0435	0.0572	0.0048	-0.0182
Upper critical value		0.1214	0.1358	0.0951	0.0709
Lower critical value		-0.0366	-0.0223	-0.0849	-0.1063
One-factor standard error					
Average value	0.0139	0.0148	0.0168	0.0139	0.0142
Average of differences		0.0009	0.0029	-0.00003	0.0003
Upper critical value		0.0035	0.0058	0.0030	0.0035
Lower critical value		-0.0018	-0.0001	-0.0030	-0.0028
Four-factor standard error					
Average value	0.0133	0.0142	0.0158	0.0133	0.0135
Average of differences		0.0009	0.0025	-0.00004	0.0002
Upper critical value		0.0033	0.0051	0.0027	0.0029
Lower critical value		-0.0016	-0.00005	-0.0027	-0.0026

This shows results from examining the risk of 10,000 bootstraps of portfolios of 100 stocks. Upper and lower critical values are the 95th and 5th percentile values of differences between subtracting unscreened from screened portfolios, and provide confidence intervals for our tests. Portfolios are value weighted. No screen represents portfolios formed from the full sample of all S&P 500 firms. Pos portfolios are formed from S&P 500 firms which KLD identifies as having strong performance across community, corporate governance, diversity, employee relations, human rights and product. Pos (ES) portfolios weight each dimension equally, whereas with Pos (WS) portfolios, each dimension is weighted according to the number of criteria it encompasses. Neg (SIC) are formed by excluding S&P 500 firms whose NAICS or SIC indicate they are from the tobacco, alcohol, gambling or weapons industries. Neg (KLD) portfolios are formed by excluding firms which KLD identifies as being involved in tobacco, alcohol, firearms, gambling, weapons, military or nuclear. The sample period is January 1996–December 2010

** Significance at the 5 % level

and UMD.¹² In this case, we get *no* significant differences across any of our performance or risk measures.

We also run our tests using the CRSP value-weighted dividend inclusive index for our market proxy, instead of the S&P 500 index. This is a less appropriate benchmark,

¹² The industry returns data are from Kenneth French's data library, as is the file which matches SIC and Fama–French industry classification codes. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html—date accessed 12 December 2011. We are grateful to Kenneth French for making these data available.

as our funds only draw stocks from the S&P 500 index, but since it is a widely used benchmark in finance research, we investigate whether our results are sensitive to choice of benchmark. We need to respecify our factors used in the four-factor model calculations. We therefore extract Fama–French size, book-to-market and momentum factors from the Fama–French Portfolios and Factors database. There is no evidence of any difference in the returns or total and systematic risk of our screened and unscreened portfolios. Unsurprisingly, all standard error results are higher across

all our portfolios: obviously, the portfolios track the S&P 500 better than the CRSP VW index. We also find that the standard error on the four-factor Pos (WS) is significantly higher than that of the unscreened portfolio. This result is, however, more likely attributable to benchmark misspecification than anything else.

In our final robustness test, we extend the investible universe of stocks beyond the S&P 500.¹³ As can be seen from Table 1, although the largest proportion of SRI funds benchmark to the S&P 500, many funds have a much broader investible universe. However, KLD data on a wider universe has only become available more recently: KLD has scored the largest US 3,000 listed stocks since 2003. We therefore rerun our analysis by comparing screened and unscreened portfolios formed from this larger universe beginning in 2003. We again do not find significant differences using any of our performance or risk measures.

Conclusion

Socially responsible investing involves removing firms which are considered to be undesirable from a portfolio, dubbed “negative screening”, or including desirable firms into a portfolio, “positive screening”. In this article, we investigate the relative impact of positive and negative screening on portfolio performance and risk. We mimic the type of portfolios retail investors are likely to hold by forming portfolios with the same characteristics as SRI mutual funds while at the same time removing the confounding issues of fees, expenses and managerial skill.

We find no evidence of either positive or negative screening impacting portfolios’ risk or returns. In very large portfolios, which are substantially larger than portfolios typical mutual funds might hold, intense screening may increase idiosyncratic risk. This finding, however, is unlikely to be relevant to the typical SRI mutual fund holder whose fund will comprise just over 100 stocks.

Our results contradict some of the inflammatory rhetoric which has surrounded SRI. Specifically, recent literature has found higher returns for sin stocks and concludes that investors who exclude these stocks will suffer dire consequences: substantially reduced returns (Adler and Kritzman 2008; Fabozzi et al. 2008). However, we uncover that there is an extremely small number of pure play sin stocks available within the S&P 500—the primary benchmark used by equity mutual funds—which makes this proposition questionable. Indeed, even using stronger negative screening which also excludes firms with *involvement* in these industries (firms which are not “pure play” sin stocks) does not result in underperformance. Negative

screening does not significantly impact on investments’ returns nor on their risks. We also do not find evidence for the assertion that positive screening will result in higher returns and lower idiosyncratic risk; at least using KLD’s data.

This is not to say that individual SRI fund managers may not out- or underperform. Our results simply suggest that any differential performance in the risk and return of individual SRI and conventional funds is not driven by the positive or negative screening utilised by the SRI funds.

In conclusion, then, we do not find evidence of either positive or negative screening impacting the return or risk of a portfolio designed to mimic a typical mutual fund’s holdings. Our results, therefore, once again uphold the now long established finding that investing socially responsibly in and of itself will not result in significant benefits or costs for investors.

Acknowledgments The authors thank Doug Foster, Lars Hassel, Steven Roberts, Paul Tan, Garry Twite, Geoff Warren, seminar participants at the Australian National University, Ecole Polytechnique, the University of Sydney Business School’s research forum on dynamics of investing responsibly and two anonymous referees for helpful comments. We thank Candice Tan for proofreading the manuscript. We also thank Chen Cheng, Kunjal Mehra, Yaokan Shen, Theingi Oo and Ying Xia for research assistance. We acknowledge the College of Business and Economics at the Australian National University for research funding.

References

- Adler, T., & Kritzman, M. (2008). The cost of socially responsible investing. *Journal of Portfolio Management*, 35(1), 52–56.
- Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, 27(11), 1101–1122.
- Bauer, R., Koedij, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance*, 29(7), 1751–1767.
- Bello, Z. Y. (2005). Socially responsible investing and portfolio diversification. *Journal of Financial Research*, 28(1), 41–57.
- Benson, K. L., Brailsford, T. J., & Humphrey, J. E. (2006). Do socially responsible fund managers really invest differently? *Journal of Business Ethics*, 65(4), 337–357.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57–82.
- Cornell, B., & Shapiro, A. C. (1987). Corporate stakeholders and corporate finance. *Financial Management*, 16(1), 5–14.
- Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, 101(3), 621–640.
- Fabozzi, F. J., Ma, K. C., & Oliphant, B. J. (2008). Sin stock returns. *Journal of Portfolio Management*, 35(1), 82–94.
- Fama, E. F., & French, K. R. (1997). Industry cost of equity. *Journal of Financial Economics*, 43(2), 153–193.
- Filbeck, G., Gorman, R., & Zhao, X. (2009). The best corporate citizens: Are they good for their shareholders? *Financial Review*, 44(2), 239–262.

¹³ We thank an anonymous referee for this suggestion.

- Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Returns and risk in socially responsible investment. *Journal of Banking & Finance*, 32(12), 2646–2654.
- Goldreyer, E., & Diltz, D. (1999). The performance of socially responsible mutual funds: Incorporating sociopolitical information in portfolio selection. *Managerial Finance*, 25(1), 23–36.
- Hong, H. G., & Kacperczyk, M. T. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1), 15–36.
- Hong, H. G., & Kostovetsky, L. (2011). Red and blue investing: Values and finance. *Journal of Financial Economics*, 103(1), 1–19.
- Humphrey, J. E., & Lee, D. D. (2011). Australian socially responsible funds—Performance, risk and screening intensity. *Journal of Business Ethics*, 102(4), 519–535.
- Jegourel, Y., & Maveyraud, S. (2010). A reassessment of the European SRI Funds “underperformance”: Does the intensity of extra-financial negative screening matter? *Economics Bulletin*, 30(1), 913–923.
- Lado, A. A., & Wilson, M. C. (1994). Human resource systems and sustained competitive advantage: A competency-based perspective. *Academy of Management Review*, 19(4), 699–727.
- Lee, D. D., Humphrey, J. E., Benson, K. L., & Ahn, J. Y. K. (2010). Socially responsible investment fund performance: The impact of screening intensity. *Accounting & Finance*, 50(2), 351–370.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.
- Renneboog, L., ter Horst, J., & Zhang, C. (2008). The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. *Journal of Corporate Finance*, 14(3), 302–322.
- Schröder, M. (2007). Is there a difference? The performance characteristics of SRI Equity indices. *Journal of Business Finance and Accounting*, 34(1–2), 331–348.
- Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal*, 56(3), 30–39.
- Statman, M. (2004). The diversification puzzle. *Financial Analysts Journal*, 60(4), 44–53.
- Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4), 33–46.
- Turban, D. B., & Greening, D. W. (1997). Corporate social performance and organizational attractiveness to prospective employees. *Academy of Management Journal*, 40(3), 658–672.
- US SIF. (2012). *Report on sustainable and responsible investing trends in the United States 2012*. Washington, DC: US SIF.