Is Socially Responsible Investing A Luxury Good?*

Ravi Bansal
Fuqua School of Business
Duke University
pll **Di (Andrew) Wu**
Stephen M. Ross School of Business
University of Michigan
Amir Yaron
The Wharton School
University of Pennsylvania

November 30, 2018

*This paper has been invited to *Revise and Resubmit* at The *Review of Financial Studies*. Yaron thanks the support from the Rodney White and Jacob Levy Centers at the University of Pennsylvania. Send correspondence to Andrew Wu at 701 Tappan Ave, R4430, Ann Arbor, MI 48109.

Is Socially Responsible Investing A Luxury Good?

Abstract

We investigate the time variability of abnormal returns from socially responsible investing. Stocks with high social responsibility ratings significantly outperform low-ranked ones during good economic times, but significantly underperform during bad economic times. Rating reductions lead to temporarily lower abnormal returns—a feature more pronounced during good times. Short-term abnormal returns after positive CSR announcements are significantly positive during good times but negative during bad times. These findings are consistent with time-varying, wealth-dependent preferences toward SRI, resulting in high-ranked stocks behaving similarly to luxury goods. Their returns are indeed significantly correlated with luxury consumption and sales growth of luxury-good retailers.

The practice of socially responsible investing (SRI) has received tremendous attention from both market participants and academic researchers in multiple disciplines. Despite a large literature in corporate strategy that studies the origin and financial impact of corporate social responsibility, relatively few finance papers examine the link between SRI and stock returns at the firm level, and existing papers seem to have divergent results. On one hand, Geczy, Stambaugh, and Levin (2005) find that under a specific belief structure, investing in SRI mutual funds incurs a significant penalty in certainty-equivalent returns compared to mutual funds without such a focus. Hong and Kacperczyk (2009) also show that "sin" firms with substantial business interests in alcohol, tobacco and gambling industries earn significantly higher alphas than comparable firms in other industries. In contrast, both Kempf and Osthoff (2007) and Statman and Glushkov (2009) find that portfolios consisting of stocks with higher corporate social responsibility (CSR) ratings have significantly higher alphas, while Edmans (2011) demonstrate that, once firms are listed in the "100 Best Companies to Work For in America" by the *Fortune* magazine, they earn higher alphas, consistent with the view of "doing well by doing good". The strategy literature also does not have a consensus on the relationship between CSR and corporate financial performance, and studies using data from different periods tend to have divergent conclusions.

This paper examines the time variability of abnormal returns from investing in socially responsible firms. We utilize the firm-level dataset on corporate social responsibility ratings, published by MSCI-ESG STATS (formerly known as the KLD database). This dataset provides annual ratings of publicly traded firms (S&P 500 and later Russell 3000 constituents) on over 100 CSR-related criteria. We aggregate the ratings into eight individual category scores, as well as one overall SR score averaged across the eight categories. In our benchmark analysis, we sort our sample into decile portfolios based on these scores, and compare the average four-factor alphas, computed using 36-month rolling estimations, of the top and bottom decile portfolios. We show that, for the period of 1993-2013, firms in the top decile portfolios ("good" firms) have marginally higher alphas than those in the bottom decile in the overall category (0.15% per month), and several individual categories such as environment and product (both at 0.32% per month).

However, the alpha differences are not persistent, but display a large variability over time. First, compared to both low- and medium-SR stocks, high-SR stocks earn much more positive alphas during good economic times, but negative alphas during bad times. We produce two ex-ante forecasts of good economic times–periods when long-term P/E ratios or GDP growth projections are in the upper half of their respective 10-year rolling distributions, as well as an ex-post indicator on whether the current period is outside NBER-designated recession periods. During these good economic periods, the Good-Bad portfolio have significantly positive four-factor alphas at around 0.24% per month. By comparison, during recessions, the alphas become negative. In contrast, during economic recessions, the Good-Bad alpha differences in most individual SR categories become statistically insignificant, or even significantly negative.

In the data, cash flow-related characteristics, such as payout and profitability margins, as well as R&D to sales ratios, do not differ significantly between the top and bottom sample firms over time. This suggests that the alpha differences between them are more likely a result of discount rate differences originating from time-varying, wealth-dependent preferences toward SRI, which would result in high-SR stocks behaving akin to luxury goods. Specifically, during good economic times, households have greater financial wealth and can consequently afford to be SRI-conscious. This drives up demand for high-SR stocks, resulting in higher realized alphas. By contrast, during bad times, households face more binding wealth constraints and therefore have to pull back on their "social consciousness" and revert to consumptions of more subsistence-like products. This reduces the demand for high-SR stocks, thereby decreasing and even reversing the alpha spreads between high- and low-SR stocks.

We find evidence consistent with this framework. We directly compare the performance of SRI with that of luxury goods consumption: In our sample period of 1993-2013, the (good-bad) alpha spread is highly correlated with per capita consumption in jewelry and watches from NIPA (correlation coefficient=0.528). By contrast, the correlation with nondurable consumption growth is 0.448.[1] In addition, similar to Aït-Sahalia, Parker, and Yogo (2004), we construct the real sales growth of a portfolio of US luxury retailers (e.g. Tiffany, Gucci, etc.) The alpha spread is also significantly correlated with real luxury sales with a coefficient of 0.329. By contrast, the cash flow characteristics, such as earnings on book equity, are not significantly different between high- and low-SR stocks during either good or bad times.

We provide further evidence for this hypothesis through long- and short-term event studies. We first examine changes in annual abnormal returns around the event of significant SR score decreases of at least two deciles.[2] We also perform this analysis separately for good and bad economic periods. We

[1]The pattern using Spearman's rank correlation coefficients is similar. The correlation with luxury consumption growth is 0.367 and that with nondurable consumption growth is 0.180.

[2]We do not use score increases in this analysis because in 2009, MSCI introduced a significant revision in ratings criteria, which

show that the negative abnormal returns are of larger magnitude, and have more statistical significance, during good economic times, but are not significant during bad times. Next, because long-run return changes might not be solely a reaction to CSR activities, we examine the short-term market reactions to positive CSR announcements. Specifically, we match 1,297 public firms in our sample to CSRWire, a voluntary press release distribution service similar to PR Newswire, but dedicated only to issues related to corporate social responsibility engagements. Compare to annually updated ratings, the advantage of this data is that firms announce CSR-related engagements continuously thorough time, thus allowing for short-term event studies on reactions specifically in response to these announcements. We examine the one-day abnormal stock returns following 5,327 CSR-related announcements, and show that on average, these announcements do not lead to significant market reactions. However, they are associated with significantly positive abnormal returns (0.09%) during good economic times and mildly significant negative returns (-0.06%) during bad times.

Our results thus indicate that the performance wedge between "good" and "bad" stocks are probably more attributable to the demand side, i.e. to the shifting of investors' SRI preferences, rather than to cash-flow-related explanations e.g. firms using SRI as advertising to differentiate their products or avoid future regulatory penalties. The luxury-good-like performance of SRI suggests that such preference shifts are related to aggregate wealth.

Our results seem surprising given the findings from papers such as Geczy et al. (2005), who show that under a particular belief structure, investing strictly in mutual funds that self-identify as SRI would realize returns that range from 0.31% to 15% lower than without such a restriction. To reconcile this difference, we first show that these funds probably use benchmarks other than MSCI ratings: despite the funds self-identifying as SRI, they did not actually allocate more toward stocks with high SR ratings *according to the MSCI definition*. Second, unconditionally, there is no statistically significant difference in alphas between the SRI and Non-SRI fund portfolios. Therefore, under the MSCI definition, the self reported SRI mutual funds are similar to other funds in their investment objectives and unconditional performances. Therefore, using the MSCI definition, we could not find discernible performance differences from mutual funds to support the alternative hypothesis of (costly) SRI initiatives. Furthermore, we conduct a series of tests to ensure the robustness of our main results. We demonstrate that our results are robust to alternative model specifications, such as choices on estimation window length, risk

resulted in many firms having discontinuous "jumps" in their scores that might not be directly attributable to CSR activities.

factor selection, and different control variables.

Our findings have several implications for investment management. First, our results suggest that SRI-focused investment vehicles might not necessarily sacrifice performance relative to market benchmarks. To illustrate this point, we eliminate short selling and construct a net long portfolio of stocks, weighted by the overall SR score and rebalanced annually. We demonstrate that, while the alphas become statistically insignificant in some specifications (particularly using equal-weighted market and quality-minus-junk risk factors), these net-long portfolios are still able to deliver performances on par with the market. Second, our findings on the luxury-good-like behavior of SRI investment portfolios highlights the importance of both good economic forecasting ability, and the ability to forecast both SRI "hot spots", in designing SRI products that deliver potentially superior performances.

Our paper is also related to the large literature in management and strategy that examines the link between corporate social responsibility and corporate financial performance. These papers usually examine whether CSR is associated with changes in firms' cash flows. Other finance studies, such as Jagannathan, Ravikumar, and Sammon (2016), suggest that CSR might be related to long-run cash flow risks. In contrast, we focus on the outcome of CSR from an investor preference perspective. Our results complement other CSR studies in finance that use richer preference structures such as Daniel, Litterman, and Wagner (2016), and highlights the importance of investor preferences in shaping the financial performance of CSR.

The rest of the paper is organized as follows: Section 1 describes our sample and data sources. Section 2 reports the results of our empirical tests. Section 3 examines the robustness of these results, and Section 4 concludes.

1 Data

1.1 Social Responsibility Ratings Data

We obtain data on social responsibility ratings from the MSCI ESG STATS (hereafter referred to as ESGSTATS) database,[3] which is a comprehensive set of annual ratings on over 100 criteria related to CSR practices by large publicly traded companies. The ratings data are published near the end of each calendar year, starting in 1991. We obtain ESGSTATS data from January 1991 to December 2013.

[3]Prior to 2011, the database was maintained by Kinder, Lydenberg, Domini & Co. and known as the KLD Social Ratings Database. Updated data are available at http://www.msci.com/products/esg/stats/

Because we use rolling 36-month regressions to compute abnormal returns, the sample period for our return-based analyses is December 1993 to December 2013. The universe of companies covered by ESGSTATS varies by time: S&P 500 companies from 1991 to 2000, Russell 1000 companies from 2001 to 2002, and Russell 3000 companies from 2003 onwards.[4]

The ratings for each firm are constructed as follows. First, eight major categories related to corporate social responsibility are identified. The categories are designated to measure firms' externalities on stakeholders in specific fields such as:

- Community: the firm's impact on the community it operates in
- Diversity: practices that affect racial and gender diversity within the firm
- Employment: labor relations, hiring practices and employee satisfaction
- Environment: the firm's environmental impact such as energy usage and carbon footprint
- Governance: investor relations, managerial transparency, and moral hazard problems
- Human Right: practices that promote or harm human rights
- Product: general product quality, and impact of the firm's products on society and environment
- Sin: concerns related to alcohol, tobacco, gambling, and pornographic industries

Within each category are a series of criteria, designated as either *strengths* or *concerns*. For each specific criterion, the MSCI conducts independent research, using both publicly available information such as news, SEC filings and legal records, and proprietary methods such as surveys and managerial interviews. It then makes a true/false evaluation for each criterion. We reproduce two criteria in the Environment category as an example:

- *Pollution Prevention (Strength): This indicator measures a firm's method of mitigating non-carbon air emissions, water discharges, and solid waste from its operations. Factors affecting this evaluation include, but are not limited to, initiatives to reduce a firm's non-carbon air emissions from its operations; to reduce the release of raw sewage, industrial chemicals, and other regulated substances; to reduce hazardous and non-hazardous waste; and programs to reduce the use of packaging materials, to support recycling; and to recycle old products such as televisions and other consumer electronics.*
- *Substantial Emissions (Concern): This indicator measures a firm's emission of toxic chemicals according to data from the Toxics Release Inventory (TRI), a U.S. Environmental Protection Agency*

[4]The Domini Social 400 index members are also included. This index has substantial overlap with the S&P 500 constituents.

(EPA) database of information on toxic chemical releases and waste management activities. Factors affecting this evaluation include, but are not limited to, how the firm compares to its industry peers.

The evaluation process results in a series of binary scores, each corresponding to a single criterion. To consolidate these ratings, we compute firm-level social responsibility scores (hereafter referred to as SR scores). For each category c, we compute the category SR score as

$$\begin{aligned} Score_t^c &= \sum_{j=1}^{s} \mathbf{I}_t^j \text{Strength}_j^c - \sum_{k=1}^{w} \mathbf{I}_t^k \text{Weakness}_k^c \\ &= \text{Number of Strengths in category } c - \text{ Number of Weaknesses in category } c, \end{aligned} \tag{1}$$

where $\mathbf{I}_t^j$ is a binary indicator that equals to 1 if a the firm exhibits a particular strength/weakness in the category. Similarly, we compute the overall SR score as the sum of individual category scores:

$$Score_t^{overall} = \sum_{c=1}^{8} Score_t^c. \tag{2}$$

Note that, while in practice some investors place higher weights on certain categories over others (for example, some SRI screens exclusively avoid sin stocks), we choose to be agnostic with respect to specific categories in computing the overall SR Score, and treat each individual SR category equally. As such, the overall SR score might be affected by categories with a larger number of criteria (e.g. Environment and Product with more than five criteria each, compared to the Sin category that has only one criterion).

[Insert Table 1 here]

We compile the top 10 and bottom 10 firms in terms of the Overall SR Score each year, and report the results from the years of 1993 and 2010 in Panel A of Table 1. We choose these two years as an example because they represent two different eras with potentially different social values and attitudes towards corporate social responsibility. We note observations: first, the top and bottom firms frequently involve well-known names, often coinciding with well-published news stories related to corporate social responsibility.[5] Second, some firms remain in top/bottom positions even after 17 years. In fact, the

[5]For example, Whole Foods Market has been known for selling organic and locally-sourced products, while McDonnell Douglas was investigated by for Department of Justice in 1993 for overcharging the government on billion-dollar military contracts. In 2010, Estee Lauder received significant media coverage for banning the use of animals in its product testing while Monsanto is alleged by the media to provide potentially harmful genetically-modified crops.

SR scores are quite persistent over time. For each firm, we compute the annual AR(1) coefficients for the overall SR score and the category SR scores, and report the mean and standard deviation of the coefficient across firms in Panel B of Table 1.[6] The scores are very persistent for the Diversity, Environment, Human Right and Sin categories, and less so for the Governance and Product categories.

1.2 CSR Announcement Data from CSRWire

In our short-term event study, we use firms' voluntary announcements of corporate social responsibility engagements through the CSRWire service. CSRWire is a press release distribution service similar to PR Newswire, but only for issues related to corporate social responsibility.[7] It has over 7,500 members including private and public firms, as well as non-profit institutions. Each member announcing CSR-related engagements can elect to have the announcement distributed by CSRWire, which sends out the press release and records the date, time, and type of each announcement. Because these engagements are voluntarily disclosed, they are most likely to be positive in nature.

Although these announcements are voluntary, they have an important advantage over annually updated ratings derived by third parties such as ESGSTATS: Firms make CSR-related announcements continuously throughout the year. Therefore, we can focus on short-term market reactions (e.g. one-day or even intraday) following these announcements, and because of the short event window, the market reactions that we observe are more likely to be responses to these announcements. This feature makes the CSRWire data suitable for testing our preference-related hypotheses.

We download 24,187 announcements from December 1999 to August 2016 using a customized web crawling algorithm. We further process these announcements to eliminate event announcements (e.g. SRI conferences and seminars) and record the identity of the announcing entity, as well as the date and time of each announcement. We then develop a fuzzy identity matching algorithm to link CSRWire firms with CRSP and Compustat Global firms, using multiple identifiers including name text, website url, and trading symbol, cross-referenced with identifiers from Worldscope and FactSet. We obtain unique GVKEYs for 1,297 firms, among which 698 firms have matched PERMNOs. These firms have 5,327 announcements in total, and these constitute our event study sample. Note that this sample covers more years than the ESGSTATS sample because of the longer coverage of CSRWire.

[6] To mitigate changes in rating criteria over time, we compute the AR(1) coefficients using the percentile rankings of the firm within a given year, rather than using the raw SR scores.

[7] See http://www.csrwire.com/ for more information.

1.3 Other Data Sources

We obtain all stock return data from CRSP and firm-level accounting data from Compustat. Table 2 reports the summary statistics of the top and bottom decile firms sorted by the overall SR score, and by each of the eight individual category SR scores. Firms in the top and bottom deciles are similar in many cash flow-related characteristics. For example, the Overall Decile 1 and Decile 10 portfolios have similar median values in book-to-market, P/E, dividend payout, and leverage ratios. Within the individual SR categories, the only noticeable differences are in terms of size: firms with high scores in Community, Employment and Human Right are on average larger than those with low scores in these categories. The top and bottom firms in other categories are similar in all characteristics including size. Because the cash flow statistics are similar for firms in the top and bottom deciles, differences in abnormal returns are potentially more likely to be a result of differences in investor preferences, rather than from the cash flow channel.

[Insert Table 2 here]

2 Empirical Tests and Results

This section presents our main results. We first examine the performances of stocks rated by overall and individual category SR scores, and compare the average monthly alphas of the top and bottom portfolios. We then examine the time-variability of the performances and link this variability with luxury good consumption data. We provide further evidence by examining long-term return changes in response to changes in SR scores, and by documenting short-term market reactions to positive CSR-related press releases.

2.1 High SR Stocks Have Higher, But Time-Varying Alphas

We first sort our sample into deciles according to the overall SR score and each of the category SR scores, and form 10 equal-weighted portfolios per score. We then construct a long-short portfolio for each category by subtracting the Decile 10 (high SR scores) portfolio returns from Decile 1 (low SR scores) returns. Next, for each month t, category c, and decile $d = 1, \cdots, 10$ (also the long-short portfolios), we fit the following regression:

$$r_{c,d,t} - r_f = \alpha_{c,d} + \beta_{c,d}^{mkt}(MKTRF_t) + \beta_{c,d}^{smb} SMB_t + \beta_{c,d}^{hml} HML_t + \beta_{c,d}^{umd} MOM_t + \epsilon_{c,d,t}, \tag{3}$$

where $r_{c,d,t}$ is the decile portfolio return in SR category c, and $MKTRF$, SMB, HML and MOM are the corresponding Fama-French and momentum factors. For each month t, we use return data from the previous 36 months to estimate the alphas and betas. Panel A of Table 3 reports, by category, the average alphas for the Decile 1 and 10 portfolios, as well as the long-short portfolio. Panel B of the same table reports the corresponding betas for the long-short portfolio. The time series of these alphas are plotted in Fig. 1, whose top panel plots monthly alphas for Decile 1 and Decile 10 portfolios separately, and the bottom panel displays the alphas of the long-short portfolio of good minus bad firms.

[Insert Figure 1 and Table 3 here]

Table 3 demonstrates a marginally significant difference in four-factor alphas between low- and high-SR portfolios in several categories. For example, Decile 1 (low SR) firms in the Environment category on average have an alpha that is 0.32% lower than that of Decile 10 (high SR) firms. Similar spreads can also be seen in the Governance (0.44% per month), Product (0.32% per month) categories and the overall score (0.15% per month). Anecdotally, the three individual SR categories that exhibit statistically significant alpha differences also constitute the bulk of investing criteria used by professional money managers.[8]

However, in the time series, Fig. 1 shows that the alpha difference between the good and bad firms is clearly *not* consistent over time.[9] For example, during the latest economic recession between 2007 and 2009, the alpha for the Decile 1 (bad) portfolio seems to be significantly higher than that of the Decile 10 (good) portfolio. This is despite the fact that the distribution of SR scores have remained relatively consistent throughout these periods. In addition, recall from Table 2 that the cash flow statistics are not significantly different between the good and bad portfolios. This suggests that the alpha differences between them are more likely to be a result of time-varying discount rate differences, which we explore in the following sections.

2.2 Economic Mechanism: Wealth-Dependent Preference Shifts

The results so far suggest that the alpha differences might originate from temporary differences in discount rates rather than from cash flow channels. This section formulates and tests a hypothesis about the source of discount rate differences over time. In particular, we argue that consumers view SRI as a

[8]According to a 2012 report by the Forum for Sustainable and Responsible Investment, four out of seven top SRI considerations for institutional investors are related to Environment, Governance, and Product categories.

[9]We obtain the same pattern for Environment, Product, Governance, and Human Right categories.

luxury good, and investing in "good" stocks are discretionary in nature. Suppose that the preference for investing in high-SR stocks (versus investing in regular stocks) is time varying and dependent on aggregate wealth: during periods when household wealth is high, households can afford to be SRI-conscious. That is, they can afford to temporarily deviate from the full-universe mean-variance efficient frontier and re-optimize based on a smaller universe consisting of high-SR stocks. This drives up demand for these "good" stocks and results in higher abnormal returns. However, during periods of low household wealth, households need to curb their discretionary spending in order to meet their basic, subsistence-based consumption. From a portfolio perspective, this means shifting back to the full mean-variance frontier and causing the demand for high-SR stocks to drop, leading to lower realized alphas.

To test this hypothesis, we construct two indicators of "good economic times" using cyclicality-adjusted real P/E (CAPE) ratios from Shiller (2005), and one-year real GDP growth projections from the Survey of Professional Forecasters. Each month between 1993 and 2013 is classified as in good economic times if either of these measures fall in the upper half of their respective 10-year rolling distributions. We first overlay these periods with the alpha time series in Figure 1. We then construct three portfolios:

- Good: Stocks with overall SR score in the top decile
- Bad: Stocks with overall SR score in the top decile
- Regular: Stocks with overall SR score *not* in the top decile

We estimate the four-factor alphas of each portfolio using 36-month rolling regressions, and report the average alphas of all three portfolios in both good and bad economic times in Table 5. This figure clearly demonstrates that the alpha difference is more pronounced during good economic times (shaded regions in Figure 1 and Column 3 in Table 5). In particular, good stocks earn much higher alphas than bad stocks during the "good times" of 1990s and 2010-2013. They earn similar alphas during "regular times" between 2003 and 2007 and significantly lower alphas than bad stocks during the financial crisis of 2007-2009.

[Insert Table 5 here]

Next, we insert the good time indicators I_t as dummies in a standard time-series regression:

$$\alpha_{p,t} = a + bI_t + \epsilon_t \tag{4}$$

where $\alpha_{p,t}$ is the (Good-Bad) or (Good-Regular) portfolio alphas. We first obtain the rolling alpha estimates by fitting Regression (3) using monthly data between 1993-2013. We then fit Regression (4) with the alpha estimates as the dependent variable. In addition to the good time dummy constructed using CAPE data, we add two more possible indicators of good economic times: a "non-recession" indicator that equals to one if the current month falls outside NBER-designated recession periods, and a continuous variable containing the (standardized) current monthly growth rate of real personal consumption expenditures. To account for any potential errors-in-variables problem resulting from the fact that our alphas are already regression estimates, we adjust all standard errors using the Shanken (1992) procedure.[10]

[Insert Table 6 here]

The results are reported in Panel A of Table 6. Again, this panel confirms the visual pattern from Figure 1. Specifically, the coefficient estimates for I_t, the good time indicator, are significantly positive in most specifications, indicating that stocks with high SR ratings do earn significantly higher abnormal returns relative to bad firms during good economic times. The excess alphas range between 0.11% to 0.36% per month, depending on the regression specification. In addition, the estimates for PCE are also mostly significantly positive, again suggesting that the demand for high-SR stocks seems higher during periods of high consumption growth. This result is consistent with our intuition that during these high-growth periods, demand for high-SR stocks might be higher due to higher aggregate wealth.

Our next tests directly compare good-minus-bad returns with characteristics of luxury goods. To do so, we first construct two series of luxury good consumption and luxury retail sales:

- Luxury PCE: NIPA-based (annual) growth in real per capita consumption expenditures in jewelry and watches.
- Luxury sales: Quarterly series of real US sales growth of five leading luxury retailers including Tiffany, Saks Fifth Avenue, Gucci, LVMH and Bulgari. This is done in a similar fashion as Aït-Sahalia et al. (2004).

We then compute the annual and quarterly average alphas of the (Good-Bad) portfolio and plot them alongside the luxury consumption and sales series, in Figure 3 below.

[10]The standard errors are generally lower if we use the Newey-West procedure without the Shanken adjustment, so we choose to report the ones with higher standard errors.

[Insert Figure 3 here]

From this figure, it is apparent that the good-minus-bad returns have high correlations with both luxury consumption and luxury retail sales. First, it is highly correlated with PCE in jewelry and watches from NIPA, with a correlation coefficient=0.528. In addition, the spread is also significantly correlated with real luxury sales with a coefficient of 0.329. This result further suggests that the performance wedge between socially good and bad stocks might be attributable to investors' time-varying preferences toward SRI, rather than to cash flow-related explanations. The luxury good-like performance of the good-minus-bad stocks suggests that such the preference toward SRI is related to aggregate wealth.

2.3 Temporary Abnormal Returns Following SR Rating Changes

The results so far suggest that stocks with high SR ratings have higher alphas than those with low ratings, during periods of high consumption growth and aggregate wealth. This is consistent with wealth-dependent preferences toward SRI. To further isolate the effect of investor preferences from other risk-related confounding factors, we employ two event studies using both long-term and short-term market reactions to possible "shocks" to firms' CSR profiles.

First, in our long-term event study, we identify all the years when the firm's overall SR score decreases by one decile or more. We start the event analysis from the end of the previous year, and compute the annual cumulative abnormal returns (CARs) as the sum of monthly alphas. We plot the time series of the average CAR from one year before to one year after the score decrease event year in Figure 2 and report the average CAR for all individual categories from one year before to five years after in Panel A of Table 4. In order to minimize any potential look-ahead bias in the alpha estimation, we also compute the average difference in raw realized returns between event firms and firms that do not experience any SR Score decrease. We report this difference in Panel B of Table 4.

[Insert Figure 2 and Table 4 here]

Both Figure 2 and Table 4 show that abnormal returns are temporarily lower after SR Score decreases. However, the magnitude is quite small and the effect short-lived. With the exception of the Employment category, none of the score categories have CARs statistically different from zero beyond $t + 2$. This again suggests that the time-varying preferences might be an important factor in driving these returns. To examine this, we perform the event study separately for good economic periods (defined by CAPE and SPF data), and report the one-year cumulative abnormal returns in Panel A of Table

7 below, which confirms that, consistent with our hypothesis, the CARs are indeed significantly negative (between -3.3% and -3.9%) during good times, but are not statistically significant outside these good times.

[Insert Table 7 here]

2.4 Event Studies Based on Voluntary CSR Disclosure

A concern for the long-term study above is that it might capture additional confounding factors beyond preference changes. We address this concern with a short-term event study of abnormal returns around firms' voluntary announcements of CSR engagements through press releases distributed by CSRWire, described in Section 1.2. For each announcement (i, t) in our sample of 698 firms and 5,327 announcements, we compute the one-day abnormal return $AR_{i,t}$ as the stock's raw return from the day of the announcement to the following day, minus the return of the CRSP value-weighted market index in the same period. We then report the average AR in both good and bad economic times (using all three indicators discussed in Section 2.2) in Panel B of Table 7 above.

Again, the result from this short-term return study is consistent with our framework of wealth-dependent preferences. First, the content of these announcements are consistent through time. Both the length and overall tone (fraction of positive words) of the announcements are not significantly different between those made during good times and those made during bad times. However, during good economic times, an announcement of CSR-related engagements is associated with a positive one-day abnormal return ranging between 9 to 11 basis points. In contrast, during recessions, similar announcements would be associated with a mildly negative return of 4 basis points. While the magnitudes of these returns are quite small, their consistent sign and statistical significance suggest that market does react differently to similar CSR announcements made during different economic climates, further lending credence to our framework of wealth-dependent preference shifts.

2.5 Practical Implications

In this subsection, we examine whether our economic insights can be applied in the practical setting of portfolio management. Specifically, our main analyses use equal-weighted long-short portfolios. This approach might require shorting a large number of small stocks. Therefore, even if our economic findings are valid, their practical implications might be limited. We address this concern by constructing an alternative, net-long portfolio based on SR Scores, and examine whether our results hold in this setting. We start from the perspective of an SRI-focused but short-constrained investor, facing the same universe as the ESGSTATS data and having access to the ratings. Then, intuitively, she would place high weights on stocks with high SR ratings, and low (or zero) weights on stocks with low ratings. We capture this with the following main weighting scheme for each stock i and year t:

$$
\begin{aligned}
w_{i,t} &= \frac{\exp(Score_{i,t}^{TOTAL})}{1+\exp(Score_{i,t}^{TOTAL})} \\
\bar{w}_{i,t} &= \frac{w_{i,t}}{\sum_{I} w_{i,t}}
\end{aligned} \tag{5}
$$

The rescaling of w into $\bar{w}$ ensures that the portfolio weights sum up to one. We also produce two alternative weighting schemes. The first alternative scheme replaces the score *levels* with *changes* in the scores from the past year, i.e. replacing $\exp(Score_{i,t}^{TOTAL})$ with $\exp(\Delta Score_{i,t}) \equiv \exp(Score_{i,t}^{TOTAL} - Score_{i,t-1}^{TOTAL})$. The third scheme further eliminates bad firms altogether, that is:

$$
w_{i,t} = \begin{cases} \dfrac{\exp(\Delta Score_{i,t})}{1+\exp(\Delta Score_{i,t})}, & \text{if } \Delta Score_{i,t} > 0 \\ 0, & \text{if } \Delta Score_{i,t} \leq 0 \end{cases}
$$

$$
\bar{w}_{i,t} = \frac{w_{i,t}}{\left(\sum_{I} w_{i,t} \middle| w_{i,t} > 0\right)} \tag{6}
$$

We form three net-long portfolios using the three weighting schemes, and compute the alphas for these portfolios by fitting Regression (3). These portfolios are rebalanced every year. Because these portfolios are now net-long and hold a large number of stocks, we use several alternative risk factors in addition to the standard four-factor model, including (1) CAPM, (2) CAPM with equal-weighted market portfolio, and (3) the four-factor model plus the quality-minus-junk factor computed per Asness, Frazzini, and Pedersen (2014). We report these alphas in Table 8 below. In Panel A of this table, we report the alphas estimated using 36-month rolling regressions. Panel B reports the alphas from the fixed-beta regressions (i.e. fitting the regression only once for the entire sample period).

[Insert Table 8 here]

These results are broadly consistent with the previous results using long-short portfolios. In a rolling regression setting, the temporary alphas persist with most model specifications and weighting schemes. The results for good and bad periods are also similar. In the single-regression, fixed-beta case, the four-factor and market alphas remain significantly positive for all three net-long portfolios, but none of the alphas are significantly different from zero if the equally-weighted market portfolio is used to calculate the MKTRF factor, nor when the quality-minus-junk is added into the four-factor setting (the results for good/bad economic times are similar to the long-short case, and the alphas are significantly positive during good times but not significant during bad times). This indicates that while the SR-weighted net long portfolios might not outperform the market in certain specifications, they can still deliver performances at least on par with the market. Therefore, SRI-focused investment vehicles might not necessarily sacrifice performance relative to market benchmarks, and our economic mechanism of wealth-dependent SRI preference also seems to hold for the net-long portfolios.

3 Other Robustness Tests

3.1 Relation to Previous Results

In contrast to our results, findings in both Hong and Kacperczyk (2009) and Geczy et al. (2005) suggest that SRI might be uniformly unattractive. First, Hong and Kacperczyk (2009) construct a portfolio of firms with substantial business interests in alcohol, tobacco and gambling industries. They find that such firms, many of which are the bottom decile firms in the Sin category of our sample, earn significantly higher alphas than comparable firms in other industries in the 1964-2004 period. In contrast, while firms in our bottom Sin decile do have higher alphas in the 1993-2011 period than those in the top decile, the difference is not statistically significant. Our different results can partly be explained by different composition of the Sin portfolios. In Hong and Kacperczyk (2009), the portfolio is selected based on NAICS/Fama-French industry codes and Compustat segment data. As a result, it consists of companies whose sole business, or the majority of its business, falls within the Sin category. Our Sin portfolio is selected using a much wider set of criteria and, as a result, includes companies not necessarily in the Sin industries/segments but conduct business related to Sin firms in some form, e.g. the supplier of ingredients used in alcohol production. Our Sin portfolio is therefore much larger than that in Hong and Kacperczyk (2009) and have potentially different return profiles depending on investor preferences over specific Sin criteria.[11]

In addition, Geczy et al. (2005) utilize a different structure where investors have prior beliefs about mutual fund managers' skills and about the accuracy of various pricing models, then update their beliefs according to fund returns generated from a pricing model that incorporates additional factors that are latent in Fama-French but co-vary with traditional Fama-French and momentum factors.[12] Next, investors form ex-ante efficient portfolios by adjusting their mutual fund holdings to maximize overall Sharpe ratios. Geczy et al. (2005) find that, under such a structure, if investors restrict their universe to only SRI funds, then those who believe in the Fama-French pricing model require a certainty-equivalent monthly return that ranges from 0.31% to 15% lower than investors without such a restriction. They report no difference in required returns for investors who believe in the CAPM. While the finding for Fama-French believers seems to support the alternative hypothesis of SRI being a net negative endeavor, our time-varying results indicate that this might not necessarily be the case using firm-level data. We conduct two tests to reconcile these results. We first obtain from CRSP quarterly data on portfolio holdings of the self-identified SRI mutual funds covered in Geczy et al. (2005) (hereafter referred to as the GSL sample), as well as holdings data from a control sample of non-SRI, no-load equity mutual funds, from 2002-2011.[13] Then, for each portfolio in the two samples, we construct the quarterly portfolio-level SR score as the holding-percentage-weighted average of SR scores of all individual stocks covered by the ESGSTAT database. We then plot the time series of mean overall SR score for the GSL and control samples in Figure 4.[14] This figure shows that the overall SR score is similar between the GSL and control samples in the 2002-2011 period. This suggests that, despite the funds self-identifying as SRI, they might not necessarily allocate more toward stocks with high SR ratings according to the ESGSTAT definition, but rather use more proprietary screens. Second, to assess the average performance of SRI and other mutual funds, we form equal-weighted portfolios of the SRI funds reported by Geczy et al.

[11] Another difference between the samples is that our Sin portfolio include firms doing business with the pornography industry, which comprise roughly 10% of the portfolio. In addition, in untabulated results, we also extend the sample in Hong and Kacperczyk (2009) from 2004 to 2011. We find a mildly higher alpha for the sin portfolio. Detailed results are available upon request.

[12] Specifically, GSL use four additional factors, which are the four principal components constructed from 20 value-weighted industry portfolios.

[13] No portfolio holdings data prior to 2002 is available. Geczy et al. (2005) report 34 mutual funds that identify themselves as SRI funds. We are able to match all of them with data on portfolio holdings. Some funds have added portfolios or classes of shares under the same investment principle after December 2001. As a result, our sample consists of 47 portfolios.

[14] We also compute the differences in mean SR Scores, by category, between GSL and control samples. The difference is not statistically significant in all but the Environment and Sin categories. In the Environment/Sin category, the GSL sample portfolio have a higher/lower SR Score than the control sample.

(2005), and form a comparison portfolio of other no-load, equity funds. In a model-free setting, we construct the long-short portfolio of (SRI funds–other funds) and report the monthly four-factor alphas and betas in Table 9. In this setting, there is no statistically significant difference in alphas between the SRI and Non-SRI fund portfolios. The results are similar when we extend the sample to 2013, as well as when we examine only the post-GSL sample period of 2002-2013. Therefore, under the ESGSTAT definition of SRI, the self-reported SRI funds are similar to other funds in their investment objectives and realized performance in a model-free setting.

[Insert Figure 4 and Table 9 here]

3.2 Alternative Specifications

We first establish the robustness of our results to alternative specifications of estimation window lengths in Table 6, where we demonstrate that the results are similar when we use rolling estimation windows from 36 to 60 months. Furtuermore, the Shanken adjustments to the standard errors in these regressions might still not completely address the error-in-variables problem associated with rolling regressions. We therefore propose another specification of time-varying betas, but with the time variation captured by indicator variables instead of rolling estimations. Specifically, we extend Equation (4) and interact the "good time" dummy I_t with the full set of covariates:

$$\begin{aligned} r_{i,t} - r_{f,t} &= \alpha_0 + \alpha_1 I_t + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t \\ &\quad + \gamma_1 (MKTRF_t \cdot I_t) + \gamma_2 (SMB_t \cdot I_t) + \gamma_3 (HML_t \cdot I_t) + \gamma_4 (MOM_t \cdot I_t) + \epsilon_{i,t}. \end{aligned} \tag{7}$$

In this setting, we essentially allow for the betas to vary according by good/bad economic times. While this does not allow for full time variability as in the rolling regressions, it can serve as a robustness check to our main results. We fit the above regression using I_t derived from CAPE, non-recession indicator, and the continuous variable of real PCE growth. We report the results in Panel B of Table 6. The four-factor alphas are still significantly positive across all specifications, and the CAPM alphas are mildly positive as well. While the magnitudes of the alphas are not directly comparable to those in Panel A due to the additional interaction terms, the fact that they remain significantly positive suggests that our results are robust to this alternative specification.

4 Conclusion

We provide a comprehensive analysis of the returns and risks of socially responsible investing, utilizing firm-level data on social responsibility ratings. We show that "good" firms with higher SR ratings have mildly higher average alphas than "bad" firms. However, these alphas are time varying, with high-ranked stocks significantly outperforming low-ranked ones during good economic times, but significantly underperforming them during bad economic times. In addition, we use two event studies to demonstrate that (1) reductions in firms' social responsibility ratings lead to temporarily lower abnormal returns, and this fact is more pronounced during good economic times, and (2) abnormal returns after positive CSR-related press announcements are significantly higher in good economic times than in bad economic times. These findings are consistent with time-varying, wealth-dependent preferences toward SRI, which could result in more socially responsible stocks behaving in a fashion akin to luxury goods. Indeed, we find that the alpha difference is significantly correlated with both luxury consumption from NIPA and sales growth of leading luxury-good retailers.

References

Aït-Sahalia, Y., J. A. Parker, and M. Yogo. 2004. Luxury goods and the equity premium. *The Journal of Finance* 59:2959–3004.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2014. Quality minus junk. *AQR Working Paper, AQR Capital Management* .

Daniel, K. D., R. B. Litterman, and G. Wagner. 2016. *Applying asset pricing theory to calibrate the price of climate risk*. Working Paper, Columbia University.

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

Geczy, C., R. F. Stambaugh, and D. Levin. 2005. *Investing in Socially Responsible Mutual Funds*. Working Paper, University of Pennsylvania.

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

Jagannathan, R., A. Ravikumar, and M. Sammon. 2016. *Environmental, Social, and Governance Criteria: Why Investors Should Care*. Working Paper, Northwestern University.

Kempf, A., and P. Osthoff. 2007. The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management* 13:908–922.

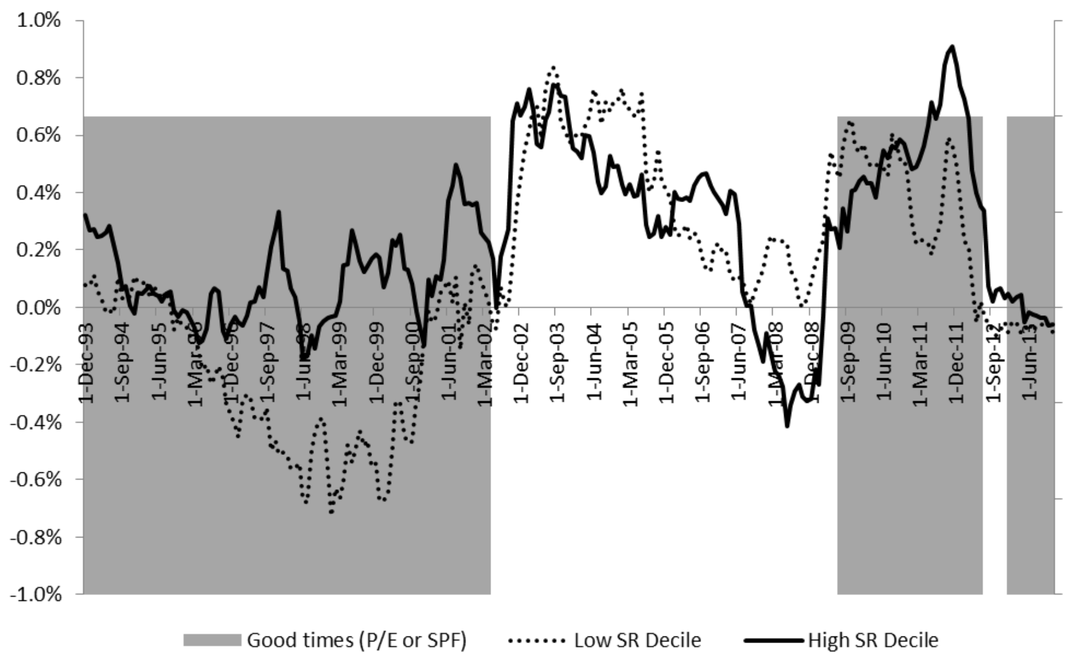
Shanken, J. 1992. On the estimation of beta-pricing models. *Review of Financial studies* 5:1–33.

Shiller, R. J. 2005. *Irrational exuberance*. Princeton University Press.

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

Figure 1. Alphas for Top and Bottom SR Decile Portfolios and the (Top-Bottom) Portfolio

The top figure displays the time series of monthly Fama-French four-factor alphas for the top- and bottom-decile portfolios sorted on the overall SR Score, defined in Eq. 2 of the text, from December 1993 to December 2013. The bottom figure displays the (Decile 10-Decile 1) portfolio alphas over the same period. Decile 1 portfolio consists of the stocks with the lowest SR ratings, and Decile 10 portfolio consists of stocks with the highest SR ratings. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013. The shaded area corresponds to "good economic times" defined using either long-term P/E ratios or GDP growth projections from Survey of Professional Forecasters. See Section 2.2 for details. All alphas are estimated using 36-month rolling regressions, where each stock is required to have at least 6 months of return data available.



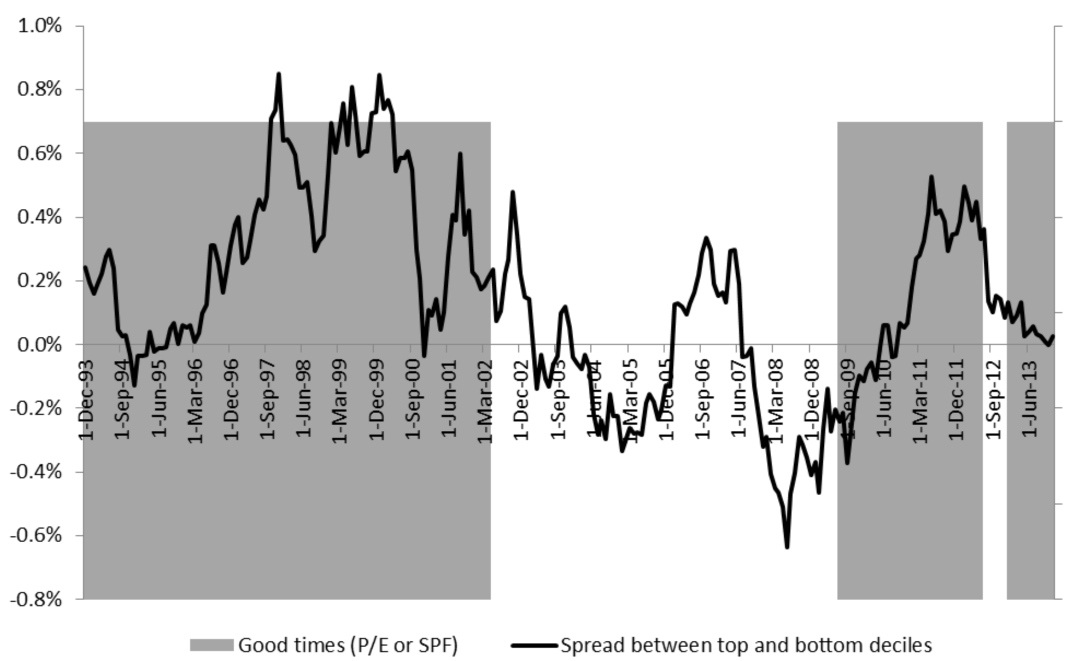


Figure 2. Time Series of Alphas Before and After SR Score Decrease

This figure displays the time series of mean monthly Fama-French four-factor alphas from one year before $(t-1)$ to one year after $(t+1)$ the event when the firm's overall SR score, computed per Eq. 2 of the text, decreases by more than one point during current year (t). The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2011. All alphas are estimated using 36-month rolling regressions with up to 217 observations per stock, and each stock is required to have at least 6 months of return data available.

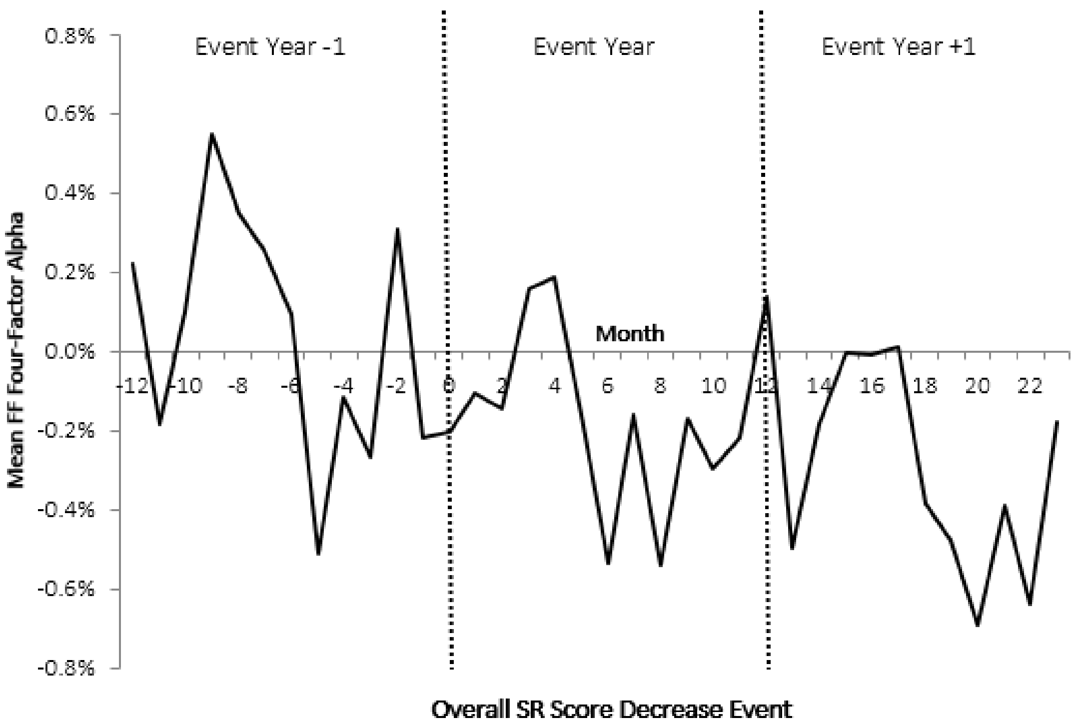
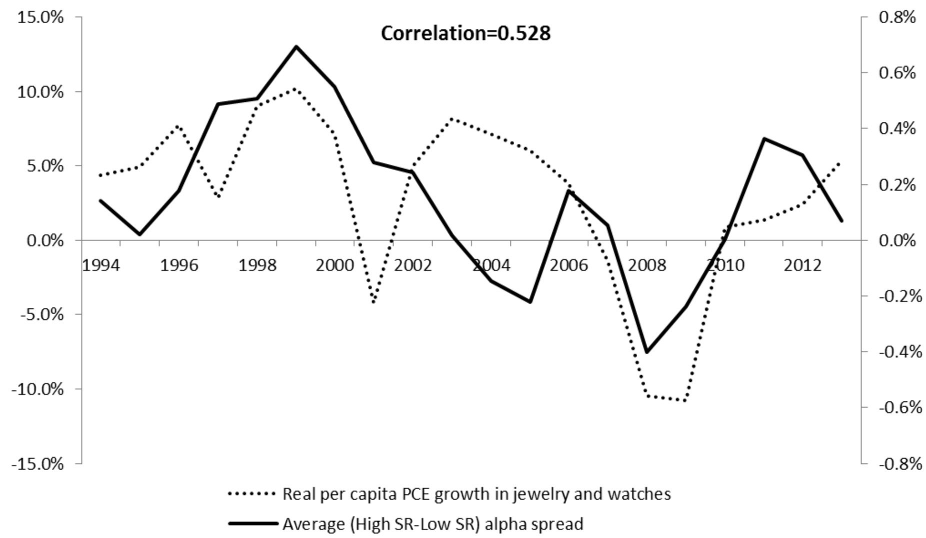


Figure 3. Correlation between (Good-Bad) Portfolio Alphas and Luxury Consumption

The top panel of this figure plots the annual (average of monthly) alphas of the (good-bad) portfolio against the annual growth rates in real per capita personal consumption expenditure growth in jewelry and watches, obtained from the National Income and Product Accounts (NIPA). The bottom panel plots the quarterly (average of monthly) alphas of the (good-bad) portfolio against the quarterly revenue growth rates of five leading luxury retailers including Tiffany, Saks Fifth Avenue, Gucci, LVMH and Bulgari, computed per Aït-Sahalia et al. (2004). The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013. All alphas are estimated using 36-month rolling regressions, where each stock is required to have at least 6 months of return data available.



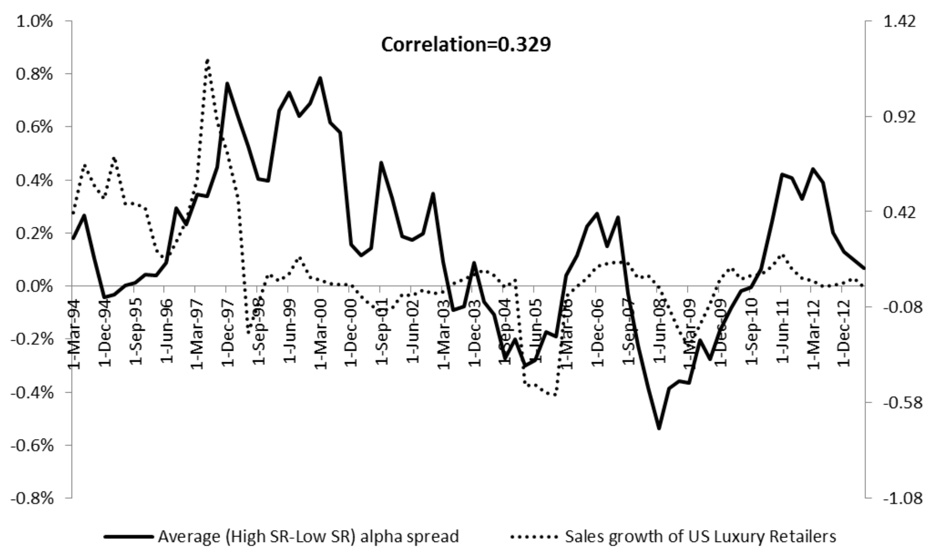


Figure 4. SR Score for GSL Fund Holdings

This figure displays the mean quarterly SR scores for the portfolio holdings of mutual funds that are self identified as socially responsible and reported in Geczy et al. (2005), as well as the mean SR score for the holdings of all other no-load equity mutual funds. The mean SR score is computed as the average of the overall SR score, defined in Eq. 2 of the text, of all stocks within the mutual fund portfolios that are within the MSCI coverage universe. Our GSL sample consists of 47 portfolios belonging to 34 socially responsible mutual funds, and our comparison (Non-GSL) sample consists of 8,453 portfolios belonging to 1,312 no-load equity mutual funds.

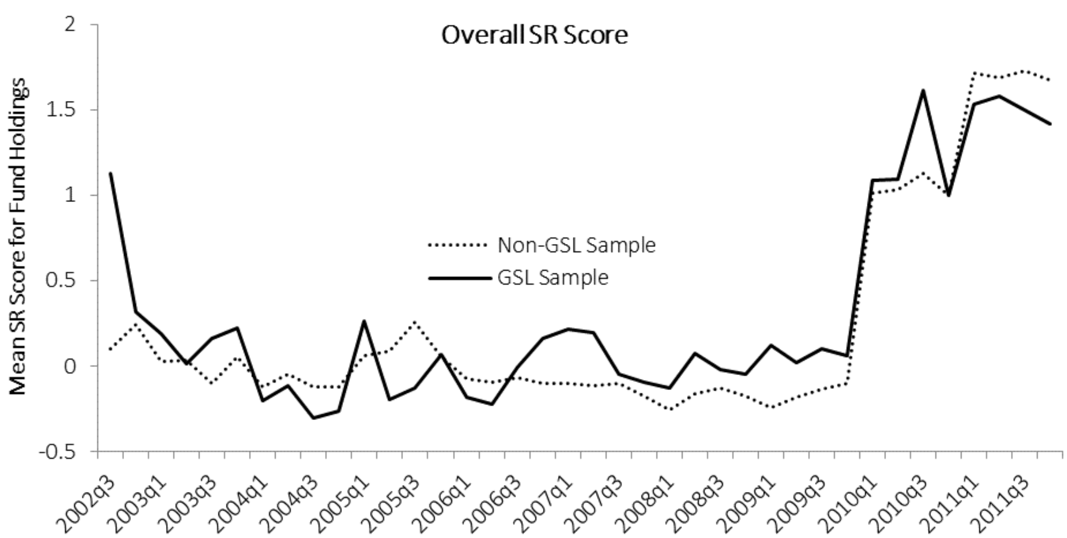


Table 1. Top and Bottom Firms in SR Scores and Persistence in Scores

Panel A of this table presents the names of the top 10 and bottom 10 firms in terms of the overall SR score, computed per Eq. 2 of the text, in the year of 1993 and 2010, respectively. Panel B of the table reports the cross-section mean and standard deviation of the AR(1) coefficient of the SR percentile scores. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2011. The AR(1) coefficients are computed once per stock, using all available score data in the sample period. The stock is excluded from the sample if there are gaps between coverage years.

Panel A: Top and Bottom 10 Firms in Overall SR Scores, 1993 and 2010

1993		2010	
Top 10	Bottom 10	Top 10	Bottom 10
Lotus Development Corp	Occidental Petroleum	Texas Instruments	L-3 Communications
Herman Miller Inc	International Paper Co	Johnson & Johnson	PulteGroup
CoreStates Financial	UNISYS Corp	Avon Products	Halliburton
Apple Computer Inc	McDonnell Douglas	Xerox Corp	Cintas Corp
Polaroid	USX Corp	AMD Corp	Stericycle Inc
Xerox Corp	Marathon Group	Procter & Gamble	Lorillard Corp
General Mills	NL Industries Inc	Starbucks	Monsanto Corp
Whole Foods Market	Phelps Dodge Corp	General Mills	KBR Inc
Ben & Jerry's	American Cyanamid Co	Northern Trust	Las Vegas Sands Corp
Sara Lee Corp	Lockheed Corp	Estee Lauder	Casella Waste Systems

Panel B: Persistence of SR Percentile Scores Across Years

Score Catetory	Cross-Section Summary AR(1) Coefficients		
	No. of Obs.	Mean	StDev
Overall	791	0.535	0.258
Community	790	0.547	0.232
Diversity	785	0.671	0.236
Employment	790	0.515	0.257
Environment	790	0.629	0.214
Governance	791	0.370	0.278
Human Right	790	0.712	0.172
Product	790	0.539	0.255
Sin	34	0.643	0.141

Table 2. Summary Statistics

This table presents the portfolio mean and median of firm-level characteristics outlined in Section 1 of the text. All mean and median values are computed as the cross-section mean/median of the time series averages of each firm in the respective portfolio, sorted on 9 SR Score categories. Decile 1 portfolio consists of the stocks with the lowest SR ratings, and Decile 10 portfolio consists of stocks with the highest SR ratings. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013.

Score Category	Score Decile	Size ($mln)		Book-to-Market		Price-to-Earning		Dividend Payout		Debt-to-Equity	
		Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Overall	1 (Low)	2887.535	281.144	0.816	0.600	22.931	12.252	0.258	0.000	0.907	0.129
	10 (High)	2183.508	235.500	0.793	0.566	14.289	12.633	0.088	0.000	0.642	0.131
Community	1 (Low)	2034.618	147.450	0.786	0.570	27.191	12.109	0.217	0.000	0.709	0.119
	10 (High)	2868.022	233.289	0.771	0.570	13.760	13.039	0.043	0.000	0.553	0.123
Diversity	1 (Low)	3033.714	273.064	0.792	0.589	24.247	12.777	0.289	0.000	1.012	0.123
	10 (High)	845.436	68.628	0.787	0.586	22.593	11.999	0.266	0.000	0.617	0.114
Employment	1 (Low)	2041.021	146.286	0.831	0.597	26.150	11.742	0.326	0.000	0.857	0.132
	10 (High)	2652.891	265.745	0.721	0.549	11.019	13.750	0.168	0.000	0.532	0.115
Environment	1 (Low)	3056.573	305.209	0.650	0.504	23.591	14.784	0.158	0.000	0.520	0.113
	10 (High)	2612.676	303.676	0.917	0.605	15.590	12.013	0.071	0.000	0.692	0.135
Governance	1 (Low)	1804.888	123.064	0.803	0.582	32.181	11.516	0.221	0.000	0.725	0.121
	10 (High)	2263.463	165.763	0.848	0.571	25.043	12.500	0.177	0.000	0.667	0.109
Human Right	1 (Low)	2275.900	172.213	0.785	0.567	30.241	12.149	0.203	0.000	0.720	0.120
	10 (High)	2631.562	378.959	0.834	0.581	17.116	14.573	0.146	0.000	0.431	0.165
Product	1 (Low)	2850.583	228.941	0.903	0.656	36.050	12.031	0.293	0.000	0.794	0.145
	10 (High)	3133.562	307.106	0.851	0.571	16.990	12.692	0.167	0.000	0.904	0.135
Sin	1 (Low)	2040.616	147.750	0.785	0.570	29.170	12.119	0.212	0.000	0.705	0.119
	10 (High)	1828.982	225.087	0.742	0.584	17.261	11.889	0.656	0.000	0.659	0.109

Table 3. Portfolio Alphas and Betas for Different Social Responsibility Rating Categories

Panel A of this table presents the mean of monthly Fama-French four-factor alphas for the top- and bottom-decile portfolios sorted on 9 SR Score categories. Panel B displays the corresponding betas for the (Low SR-High SR) portfolio. Decile 1 portfolio consists of the stocks with the lowest SR ratings, and Decile 10 portfolio consists of stocks with the highest SR ratings. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013. All alphas and betas are estimated using 36-month rolling regressions with up to 217 observations per stock, and each stock is required to have at least 6 months of return data available. The numbers in brackets are t-statistics.

Panel A: Alphas for Decile and Spread Portfolios

Score Decile	SR Score Category								
	(1) Com	(2) Div	(3) Emp	(4) Env	(5) Gov	(6) Hum	(7) Pro	(8) Sin	(9) Overall
1 (Low)	0.00269***	0.00304***	0.00261***	0.00079	0.00236***	0.00286***	0.00136***	0.00268***	0.00196*
	(3.59)	(3.39)	(3.69)	(1.21)	(3.94)	(3.50)	(3.49)	(3.91)	(2.09)
10 (High)	0.00223***	0.00201***	0.00282***	0.00394**	0.00672***	0.00416	0.00453***	0.00181	0.00350***
	(4.65)	(4.06)	(5.61)	(3.22)	(4.77)	(1.04)	(4.46)	(1.20)	(5.58)
Low-High	0.00045	0.00104	-0.00020	-0.00315*	-0.00436**	-0.00130	-0.00318**	0.00087	-0.00154**
	(0.57)	(1.28)	(-0.40)	(-2.18)	(-3.32)	(-0.31)	(-2.88)	(0.81)	(-2.68)

Panel B: Betas for the (Low-High) Portfolio

β	SR Score Category								
	(1) Com	(2) Div	(3) Emp	(4) Env	(5) Gov	(6) Hum	(7) Pro	(8) Sin	(9) Overall
Mkt	0.151***	0.138***	0.0721**	0.00686	0.122**	-0.182	-0.0948***	0.0318	0.150***
	(9.53)	(7.20)	(2.78)	(0.17)	(2.83)	(-1.94)	(-4.30)	(0.80)	(6.16)
SmB	0.409***	0.445***	0.107*	0.0412	-0.0166	0.512**	-0.123***	0.0293	0.0852**
	(14.94)	(11.08)	(2.33)	(0.88)	(-0.18)	(2.68)	(-3.65)	(1.10)	(2.90)
HmL	-0.0470	0.0445	0.144***	0.234***	0.183	0.0129	0.512***	-0.0143	0.222**
	(-1.16)	(1.15)	(4.86)	(5.84)	(1.82)	(0.07)	(6.28)	(-0.34)	(3.29)
UmD	0.0176	0.0470	-0.0737**	-0.107**	-0.0941	0.0146	0.00480	0.0781	0.0101
	(0.78)	(1.47)	(-3.23)	(-2.74)	(-1.47)	(0.14)	(0.09)	(1.39)	(0.36)
No. Obs	217	217	217	217	217	203	217	217	217

Table 4. Cumulative Abnormal Returns Before and After SR Score Decreases

Panel A of this table reports the mean annual cumulative abnormal returns (CARs) from one year before ($t-1$) to five years after ($t+5$) the event when the firm's SR score in a given category decreases by more than one decile during current year (t). Panel B of this table reports the difference between the mean raw annual returns of all firms whose overall SR score decreases by more than one decile, and that of all other firms, during the same six-year span. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013. The CARs are computed using monthly alphas per Eq. (4) of the text. All alphas are estimated using 36-month rolling regressions with up to 241 observations per stock, and each stock is required to have at least 6 months of return data available. The numbers in brackets are t-statistics.

Score Category	Annual CAR						
	$t-1$	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Overall	-0.00171	-0.0219***	-0.0321***	-0.0261***	-0.00736	-0.0143	-0.00089*
	(-0.54)	(-6.79)	(-5.63)	(-4.43)	(-1.39)	(-1.16)	(-2.23)
Community	0.00146	0.00170	-0.0690***	0.00414	-0.0213	-0.0233	-0.00667
	(0.22)	(0.26)	(-4.81)	(0.34)	(-1.82)	(-1.95)	(-0.66)
Diversity	0.00822	-0.0207***	-0.0273**	-0.0289**	-0.0145	-0.00663	-0.0172**
	(1.54)	(-4.14)	(-3.07)	(-3.12)	(-1.72)	(-1.12)	(-3.04)
Employment	-0.00887	-0.0153**	-0.0144	-0.0359***	-0.0249**	0.00283	-0.000314
	(-1.77)	(-3.02)	(-1.57)	(-3.62)	(-2.75)	(0.33)	(-0.04)
Environment	-0.00377	-0.0235***	-0.0131*	-0.0112	-0.0289	-0.0264*	-0.0130
	(-0.55)	(-3.50)	(-2.00)	(-0.79)	(-1.27)	(-2.41)	(-1.21)
Governance	-0.00687	-0.0271***	-0.0296***	-0.0247***	0.00801	-0.0106	-0.00461
	(-1.79)	(-6.60)	(-4.41)	(-3.46)	(1.22)	(-1.88)	(-0.93)
Human Right	0.0228	-0.0198	-0.0193**	-0.0270	-0.0661	-0.0111	-0.0135
	(1.69)	(-1.20)	(-3.25)	(-1.15)	(-1.51)	(-0.45)	(-0.57)
Product	-0.00473	-0.0136*	-0.0042*	-0.00888	-0.0138	-0.00160	-0.00924
	(-0.66)	(-2.06)	(-2.35)	(-0.71)	(-1.10)	(-0.15)	(-0.90)
Sin	-0.0543	0.0210	0.0668	0.0135	0.0603	-0.0108	-0.00876
	(-1.40)	(0.62)	(1.40)	(0.32)	(1.26)	(-0.49)	(-0.53)
Panel B: Raw Return Difference (Avg. Event Returns-Avg. Nonevent Returns in the Same Period)							
Overall	0.02230	-0.03610	-0.02883	-0.00310	0.01500	-0.00170	-0.00800

Table 5. Average Alphas in Good and Bad Economic Times

Panels A and B of this table reports the mean of monthly Fama-French four-factor alphas for the top- and bottom-decile portfolios, as well as the (top-bottom) portfolio, during good and normal times defined using either long-term P/E ratios or GDP growth projections from Survey of Professional Forecasters, respectively. See Section 2.2 for details. Panel C reports coefficient estimates from Regression 4 of the text. Decile 1 portfolio consists of the stocks with the lowest SR ratings, and Decile 10 portfolio consists of stocks with the highest SR ratings. The sample consists of S&P 500 component stocks from 1993 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2013. All alphas and betas are estimated using 36-month rolling regressions, and each stock is required to have at least 6 months of return data available. The numbers in brackets are t-statistics.

Panel A: Average Alphas in Good time defined by Shiller P/E						
		Good Time			Normal Time	
	Bad (Decile 1)	Good (Decile 10)	Diff	Bad	Good	Diff
Alpha	-0.00227***	0.000888***	0.00315***	0.00345***	0.00343***	-0.000026
	(-2.85)	(2.51)	(4.12)	(5.08)	(4.15)	(-0.04)
		Good Time			Normal Time	
	Rest (Decile 1-9)	Good (Decile 10)	Diff	Rest (Decile 1-9)	Good	Diff
Alpha	-0.000682	0.000888*	0.00157**	0.00425***	0.00343***	-0.000822
	(-1.08)	(2.51)	(2.88)	(5.50)	(4.15)	(-1.35)
No. Obs	108	108	108	133	133	133
Panel B: Average Alphas in Good time defined by SPF						
		Good Time			Normal Time	
	Bad (Decile 1)	Good (Decile 10)	Diff	Bad	Good	Diff
Alpha	0.00304***	0.00416***	0.00112***	0.000312	0.00179***	0.00147
	(3.07)	(3.69)	(3.27)	(0.33)	(3.13)	(1.95)
		Good Time			Normal Time	
	Rest (Decile 1-9)	Good (Decile 10)	Diff	Rest (Decile 1-9)	Good	Diff
Alpha	0.00341***	0.00416***	0.000748*	0.00167	0.00179**	0.000116
	(2.89)	(3.66)	(2.18)	(1.86)	(3.13)	(0.20)
No. Obs	61	61	61	180	180	180

Table 6. Excess Alphas During Good Economic Times

Panel A of this table reports the coefficient estimate for I_t in Regression (4) of the text. The dependent variable, $\hat{\alpha}$, is the estimated four-factor alphas for the long-short portfolio of high SR-low SR stocks. The alphas are estimated using 36-month rolling regressions with up to 241 observations. The independent variable, I_t, is an indicator of good economic times constructed from either long-term P/E ratios or an indicator of the current month falling outside the NBER-designated recession periods. The third column of Panel A replaces the indicator I_t with the continuous variable PCE_t, which is the growth of real personal consumption expenditures from NIPA. See Section 2.2 for details. We vary the estimation window from 24 months to 60 months, and report the standard errors adjusted with the Shanken (1992) procedure. Panel B presents an alternative regression model, which serves as a robustness check. The coefficients reported are estimates of α_1 from Regression (7) of the text. The standard errors in this panel are calculated using the Newey-West procedure with 36 lags. Numbers in brackets are t-statistics.

Panel A. Rolling Beta Model: $\hat{\alpha}_{p,t} = a + bI_t + \epsilon_t$

Estimation Window	Good Time Indicator		PCE
$\hat{b}$	Shiller	Non-Recession	
36 Months	0.0025***	0.0020*	0.0069*
	(4.92)	*(2.09)*	*(2.40)*
24 Months	0.0026***	0.0033*	0.0031*
	(6.99)	*(2.18)*	*(2.34)*
48 Months	0.0030***	0.0011	0.0030
	(3.18)	*(1.63)*	*(1.40)*
60 Months	0.0036**	0.0013	0.0065*
	(2.73)	*(1.83)*	*(2.07)*

Panel B. Alternative Time-Varying Beta Model: Regression

Estimation Window	Good Time Indicator		PCE
$\hat{\alpha}_1$	Shiller	Non-Recession	
FF4	0.0047*	-0.0050	0.0487*
	(2.06)	(-0.96)	(2.04)
Market	0.0050	-0.0010	0.0294
	(1.87)	(-0.18)	(0.94)

Table 7. Market Reactions to SR Score Declines and CSR-Related Press Releases

Panel A of this table reports the mean annual cumulative abnormal returns (CARs) from one year before $(t-1)$ to one after $(t+1)$ the event when the firm's SR score in a given category decreases by more than one decile during current year (t). We compute the CARs separately according to whether the decline has occured during good economic times, defined in Section 2.2, or during recessions. Panel B reports the one-day abnormal returns (AR) after an individual firm publicly announces a CSR-related engagement thorough thr CSRWire distribution service. We compute the average over 5,327 announcements by 1,297 firms from 1999 to 2016.

Panel A: Long Term Event Study w/ KLD Scores

Event: KLD Score Decline>=2 Decile, Annual Frequency

	Annual Abnormal Returns			
	Good Time		Bad Time (Recession)	
	t-1→t	t→t+1	t-1→t	t→t+1
Shiller	-0.0138	-0.0391*		
	(-1.75)	(-2.16)		
SPF	-0.0170	-0.0334**		
	(-1.73)	(-2.80)		
Non-Recession	-0.0204*	-0.0339***	-0.0071	0.0074
	(-2.04)	(-4.98)	(-0.60)	(0.82)

Panel B: Daily Event Study Using CSR Wire Announcements

Event: Press announcement through CSRWire

	Daily Abnormal Returns			
	Good Time		Bad Time (Recession)	
	t-1→t	t→t+1	t-1→t	t→t+1
Shiller	0.0006	0.0009**		
	(0.13)	(2.74)		
SPF	0.0002	0.0010*		
	(1.33)	(2.05)		
Non-Recession	-0.0042	0.0011*	-0.0015	-0.0004*
	(-0.14)	(2.28)	(-0.82)	(-2.22)

Table 8. Alphas of SR Score-Weighted Net Long Portfolios

This table reports the alphas from rolling and fixed-beta regressions, using three net-long SR portfolios instead of the long-short portfolio. We construct these portfolios by weighting individual stocks annually by their SR Score levels, changes from the last year, and only positive changes. See Section 2.5 for details. Panel A reports the average alphas estimated using rolling regressions of various windows, with up to 241 observations per portfolio. Panel B reports the fixed alpha (intercept) from fitting Regression (3) once with the full sample. The numbers in brackets are t-statistics.

Panel A: Rolling Betas (Estimation window=36 months)

Alphas	Weighting Method		
	Level	Change	Pos Change
FF4	0.0027*	0.0019**	0.0038***
	(2.21)	(2.84)	(4.20)
Market	0.0034*	0.0029	0.0044**
	(2.05)	(1.90)	(2.83)
Equal-Weighted Market	0.0022*	0.0018	0.0034**
	(2.48)	(1.83)	(2.82)
FF4+QMJ	0.0022*	0.0018**	0.0038***
	(2.31)	(2.73)	(4.01)

Panel B: Fixed Betas

Alphas	Weighting Method		
	Level	Change	Pos Change
FF4	0.0026**	0.0018*	0.0034***
	(3.31)	(2.13)	(3.87)
Market	0.0033**	0.0030*	0.0045**
	(2.66)	(2.15)	(3.17)
Equal-Weighted Market	0.0022	0.0019	0.0035*
	(1.43)	(1.18)	(2.08)
FF4+QMJ	0.0007	-0.0001	0.0019
	(0.96)	(-0.08)	(1.94)

Table 9. Differences in Alphas Between Socially Responsible and Other Mutual Funds
This table presents the four-factor alphas and betas of the long-short portfolio that buys an equally-weighted portfolio of self-reported socially responsible mutual funds, studied in Geczy et al. (2005), and sells an equally-weighted portfolio of no-load equity mutual funds without such a focus, in five different sample periods. Column (3) corresponds to the Geczy et al. (2005) sample period. Our GSL sample consists of 34 socially responsible mutual funds, and our comparison (Non-GSL) sample consists 1,312 no-load equity mutual funds.

Portfolio: EW (GSL-Non GSL)	Sample Periods				
	(1) 1963-2013	(2) 1993-2013	(3) 1963-2001	(4) 1993-2001	(5) 2002-2013
Alpha	-0.000158 (-0.24)	0.0000866 (0.24)	-0.0000903 (-0.10)	0.000923 (1.28)	-0.000468 (-1.55)
β					
Mkt	0.0163 (1.07)	0.0504*** (5.70)	0.0131 (0.62)	0.0364 (1.81)	0.0229** (2.86)
SmB	-0.0938*** (-4.40)	-0.0568*** (-4.95)	-0.113*** (-4.15)	-0.112*** (-5.67)	0.0246 (1.79)
HmL	0.00926 (0.40)	-0.0914*** (-7.52)	0.00956 (0.30)	-0.154*** (-5.90)	-0.0565*** (-4.38)
UmD	0.0208 (1.42)	0.0237** (3.25)	0.0185 (0.88)	0.0145 (1.11)	0.0271*** (4.13)
No. of Obs	522	248	378	104	144