Socially Responsible Investing: Good Is Good, Bad is Bad

Ravi Bansal

Di Wu

Amir Yaron*

This Draft: January 10, 2016

ABSTRACT

This paper provides a comprehensive analysis of risks and returns of socially responsible investing (SRI) utilizing firm-level data on corporate social responsibility ratings. We demonstrate that firms with high ratings have significantly higher, albeit temporary and time varying, alphas than those with low ratings. In an event study setting, we find that reductions in firms' social responsibility ratings lead to significantly lower cumulative abnormal returns that dissipate after the second year. We then provide evidence indicating that these differences are induced by time-varying, wealth-dependent shocks to investors' preferences, which result in highly rated stocks behaving in a fashion akin to luxury goods. The alpha difference between stocks with high and low ratings are significantly more pronounced during good economic times, and is significantly correlated with both luxury consumption from NIPA and the sales growth of luxury-good retailers.

*Ravi Bansal is the J. B. Fuqua Professor at the Fuqua School of Business, Duke University, and NBER, email: ravi.bansal@duke.edu. Di (Andrew) Wu is a Ph.D. Student of Finance at The Wharton School, University of Pennsylvania, email: wudi1@wharton.upenn.edu. Amir Yaron is the Robert Morris Professor of Banking and Professor of Finance at the Wharton School, University of Pennsylvania, and NBER, email: yarona@wharton.upenn.edu.

1. Introduction

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 sociology and business ethics that studies the origin and proliferation of SRI, few papers in finance examine the link between SRI and stock returns at the firm level, with existing papers having seemingly 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.

This paper provides a comprehensive analysis of the returns and risks of investing in socially responsible firms. We utilize a unique 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 large publicly traded firms on over 100 CSR-related criteria. To provide a framework for our analysis, we first develop three distinct scenarios governing the behavior of realized returns of stocks with different levels of social responsibility. First, we argue that unexpected positive shocks to investors' preferences toward SRI lead to higher short-term realized alphas for stocks with higher SR ratings. By contrast, unexpected firm expenditures in SR-related activities might divert resources away from shareholders, and are thus associated with lower alphas. Finally, if SRI represents only inconsequential marketing gimmicks or measurement errors, then it would not be significantly related to any alpha difference.

We analyze our hypotheses using two approaches. In both cases, we use the firm-level ratings data and aggregate the ratings into eight individual Category SR Scores, as well as one Overall SR Score, then perform the following two tests: First, we sort our sample into decile portfolios based on the SR Scores, and compare the average monthly alphas of the top and bottom portfolios. Second, for each stock in our sample, we examine the change in returns in response to changes in SR Scores by comparing the cumulative abnormal returns before and after SR Score decreases of two or more deciles. Our first test provides support for our first hypothesis: (1) the Fama-French four-factor alphas are significantly higher for firms in the top decile portfolios ("good" firms) than those in the bottom decile, ranging from 0.15% per month (Overall SR) to 0.44% per month (Governance), (2) The alphas are temporary and time-varying. While overall high-SR firms have higher alphas, the focus of SRI with respect to specific sub-topics evolves through time: the average alpha difference between top and bottom decile portfolios in the Environment category, for example, is significantly positive prior to 2003, but dissipates afterwards. Precisely around the same time, the alpha difference in the Governance category changes from zero to significantly positive. Even at the overall level, the alpha difference reverses during bad economic times such as the recent financial crisis. Our second test then shows that, after reductions in SR Scores, average abnormal returns are significantly negative in the following year. However, the negative alphas again dissipate after the second year.

Having established the main empirical facts that high-SR stocks exhibit higher, albeit temporary and time-varying, alphas, we explore the economic mechanism that could give rise to these patterns. In particular, we argue that temporary, wealth-dependent preference shocks result in high-SR stocks behaving in a fashion akin to luxury goods. Suppose that high- and low-SR stocks are similar in all cash flow-related aspects. 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 investments that deliver the best risk-adjusted returns. This reduces the demand for high-SR stocks, thereby decreasing and even reversing the alpha spreads between high- and low-SR stocks.

Our next set of tests provides evidence consistent with this mechanism. First, 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. We demonstrate that, compared to both low- and medium-SR stocks, high-SR stocks earn much higher alphas during good times, and much lower alphas during bad times. We then 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). 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.

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 discernable performance differences from mutual funds to support the alternative hypothesis of (costly) SRI initiatives.

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. We rationalize these arguments in the upcoming version of the paper featuring a model with wealth-dependent investor preferences.

Our findings have several important implications for investment management and corporate governance. First, we demonstrate that, contrary to popular beliefs, at the individual stock level, it is possible to "do well while doing good" by investing in high-SR stocks in multiple categories and realizing higher alphas. This suggests additional growth room for well-designed, differentiated SRI-focused investment products. Second, our hypothesis that the higher alphas are due to temporary demand shocks suggest that in order for the out-performance by high-SR firms to persist, the definition of CSR should continuously evolve and new areas of CSR "hot spots" need to be discovered. Our multi-faceted Scores encompass a wide range of SR categories and therefore can potentially be an appropriate instrument to detect such changes. Third, because SRI alphas behave like luxury goods, SRI investing is likely to be very cyclical. Our results provide indications of critical turning points where the investors' preferences are likely to shift, thus provide guidance on the optimal timing of CSR-related expenditures and marketing activities.

The rest of the paper is organized as follows: Section 2 describes our sample and data sources. Section 3.1 develops our main hypotheses. Section 3 reports the results of our empirical tests. Section 5 concludes.

2. Data

2.1. Social Responsibility Ratings Data

We obtain our data on social responsibility ratings from the MSCI ESG STATS (hereafter referred to as ESGSTATS) database,[1] which is a comprehensive, annual data set of ratings on more than 100 criteria related to corporate social responsibility practices by large publicly traded companies. The ratings data are published near the end of each calendar year, starting in 1991. Our sample uses 20 years of data from 1991 to 2011. 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.[2]

The social responsibility 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 the firm's externality on stakeholders other than regular shareholders in a specific field, such as:

- Community: the firm's relation to and its impact on the community
- Diversity: practices that effects racial and gender diversity in firm and the community
- Employment: labor relations, hiring practices and employee satisfaction
- Environment: the firm's environmental impact such as energy use or carbon footprint
- Governance: investor relations, transparency in management, 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

[1] Prior to 2011, the database was maintained by KLD and known as the KLD Social Ratings Database. Available at http://www.msci.com/products/esg/stats/, and also at Wharton Research Data Services

[2] The Domini Social 400 index members are also included, and this has substantial overlap with the S&P 500 constituents.

- 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 lawsuit records, as well as proprietary method 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 (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 further consolidate the ratings, we compute firm-level Social Responsibility Scores (SR Scores hereafter). For each category c listed in the above paragraph, we compute the Category SR Score as

$$\begin{aligned} Score_t^c &= \sum_{j=1}^{s} \mathbf{I}_t^j \text{Strength}_j^c - \sum_{k=1}^{w} \mathbf{I}_t^k \text{Weakness}_k^c \\ &= \text{No. of Strengths in category } c - \text{ No. of Weaknesses in category } c, \end{aligned} \tag{1}$$

where $\mathbf{I}_t^j$ is a binary indicator that equals to 1 if a the firm exhibits a particular strength/weakness in the social responsibility 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 more weight on certain categories over others (e.g. and exclusively avoiding sin stocks), we choose to be agnostic with respect to specific categories in computing the Overall SR Score, treating each individual SR category equally. As such, the Overall SR Score will be affected by "big" categories with a larger number of criteria (e.g. Environment and Product, compared to the Sin category which has only one criterion).

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 different values and attitudes towards corporate social responsibility. Two observations are in order: first, the top and bottom firms frequently involve household names, often coinciding with well published cases related to corporate social responsibility.[3] Second, some firms remain in top/bottom spots despite the two sample years being 17 years apart. In fact, the 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.[4] The Scores are very persistent for the Diversity, Environment, Human Right and Sin categories, and less so for the Governance and Product categories. This finding suggests the interaction of two effects: first, many firms are consistent

[3]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 is commented for banning the use of animals in its product testing while Monsanto is alleged by media to provide potentially harmful genetically-modified crops.

[4]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.

in their corporate social responsibility practices. Second, preference toward SRI probably has shifted over time. The second effect has important asset pricing implications, which we further explore in Section 3.1.

2.2. 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 the eight individual Category SR Scores. Firms in the top and bottom deciles are similar in most pricing- and cash flow-related characteristics. For example, the Overall Decile 1 and Decile 10 portfolios have similar values in size, 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, any potential difference in return characteristics (e.g. alphas) is more likely to be a result from changes in preferences, rather than from the cash flow channel.

3. Empirical Facts: Alpha Differences between High and Low-SR Stocks

3.1. Hypothesis Development: Why Would Different SR Stocks Have Different Alphas?

This section formulates our main hypotheses regarding the alphas of stocks with different SR ratings. First, differences in alphas can emanate from unexpected shocks to investor's preferences. Anecdotal evidence suggests that investors have become more receptive toward CSR practices that have positive externality on other stakeholders.[5] In addition, consumers have

[5]For example, the US Forum for Sustainable and Responsible Investment (USSIF) reports that as of 2012, the total assets invested according to socially responsible criteria is $3.74 trillion, or roughly 11% of the total assets under professional management, compared to $639 billion/9% in 1995.

become more conscientious about purchasing products with positive social and environmental impact.[6] If such demand shocks are not expected by investors ex-ante, this would lead to higher realized stock returns for firms with higher SR ratings. In the long term, as the preference shocks become known by the market, the difference in returns will converge toward zero. However, the short-term mispricing could persist if there is a sequence of positive preference shocks over time (e.g. to different categories of SR), leading to higher realized stock returns for firms with high overall SR ratings. We summarize our reasoning in the following hypothesis:

Hypothesis 1 *In the presence of unexpected preference shocks that favor high-SR firms, the alphas of such firms will be higher than those of low-SR firms. The difference in alphas could persist with a sequence of such unexpected preference shocks, before eventually converge toward zero.*

An alternative hypothesis is that, if a firm takes actions to enhance its corporate social responsibility, it may require higher costs than comparable firms without such activities. Alternatively, Cheng, Hong, and Shue (2014) demonstrate that firms might shift resources away from shareholders to other stakeholders. As such, CSR is potentially a costly endeavor. If the cost is not expected by investors ex-ante, the realized returns will be lower in the short term for firms that engage in such activity. In the long term, again as the market learns of such shocks, the expected return differences will adjust toward zero. This prediction is partly reflected by findings in Hong and Kacperczyk (2009), and results in the following hypothesis:

Hypothesis 2 *In the presence of unexpected costly expenditures undertaken to enhance corporate social responsibility, the alphas of high SR firms will be lower than those of low-SR firms. The difference in alphas converges toward zero in the long term.*

It could also be argued that CSR and SRI practices are only fads or marketing strategies

[6]For example, the market share of hybrid vehicles in the US has grown from 0.06% in 2010 to 3.20% in 2013, according to Department of Energy statistics. Also according to industry data, the market share of eggs produced in a cage-free environment has grown from 1.5% in 2000 to over 5% in 2013, despite higher shelf prices.

designed to garner media attention, and have no real impact on either stakeholders or the society as a whole. Therefore, rational consumers and market participants pay no attention to these activities regardless of their preferences toward "true" SRI. In addition, the SR Scores that we use could contain large measurement errors and, in extreme cases, do not measure true CSR activities at all. In both cases, the realized returns of firms with different SR Scores would be similar.

Hypothesis 3 *In the absence of either unexpected preference shocks or unexpected SR expenditures, or if the SR Scores contain large measurement errors, then the alphas of high SR firms would be the same as those of low-SR firms.*

Ex-ante, the *expected returns* are the same for high- and low-SR stocks under all three scenarios above, despite the differences in realized returns. We now formulate a framework where the expected returns are different. Suppose that, firms' production functions contain stochastic shocks related to SR activities. An example of such shock would be uncertain regulatory actions faced by producers of alternative energy, or increased output volatility due to product sourcing from low-emission suppliers. Because such shocks are captured by the production function, if they are recognized by market participants, they represent a source of risk. High-SR firms therefore earn higher risk premia as compensation for exposure to such SR-related risks. As a result, the expected returns for high-SR firms will be higher than their low-SR counterparts. In addition, if researchers cannot measure such stochastic shocks directly from the data, the higher expected returns will translate to higher alphas in traditional asset pricing models, irrespective of whether unexpected shocks are present.

Hypothesis 4 *If CSR activities are associated with stochastic shocks to the firms' production, and market participants recognize these shocks, then firms with high SR ratings will command higher expected returns than low SR firms. These larger returns for high SR firms will appear as positive alphas if researchers do not control for such cost risk factors.*

It is worth noting that, even though both Hypotheses 1 and 4 prescribe higher alphas for high-SR stocks, the persistence these alphas are different. Under Hypothesis 1, the alphas are *temporary* and converges to zero in the long run. Under Hypothesis 4, because the "alphas" are essentially mismeasured betas, they are permanent. Thus, we can disentangle these two different sources of alphas by explicitly identifying temporary preference shocks and trying to isolate the permanence of the alphas with respect to such shocks.[7]

We test the above hypotheses using two different approaches. First, we sort our sample into decile portfolios based on the Overall and Category SR Scores, and compare the average monthly alphas of the top and bottom portfolios. Next, for each stock in our sample, we examine the change in returns in response to changes in SR Scores by comparing the cumulative abnormal returns before and after SR Score decreases.

3.2. High SR Stocks Have Higher Alphas

For each Category SR Score and the Overall Score, we sort our sample into deciles according to the Score and form 10 equally-weighted portfolios per category. We also construct a long-short portfolio (1-10) for each category by substrating 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 rolling four-factor regression:

$$r_{c,d} - r_f = \alpha_{c,d,t} + \beta^{mkt}_{c,d,t}(MKT - r_f) + \beta^{smb}_{c,d,t} SMB + \beta^{hml}_{c,d,t} HML + \beta^{umd}_{c,d,t} MOM + \epsilon, \qquad (3)$$

where $r_{c,d}$ is the decile portfolio return and MKT, SMB, HML and MOM are the corresponding Fama-French and momentum factors. For each month t, we use return and factor data from the previous 36 months to estimate the alphas and betas. Panel A of Table 3 reports, by SR Score 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

[7] In on going work, we incorporate textual analysis-based methods that identify key events related to different SR categories at the firm level and then examine the persistence in alphas.

in each category. We further explore the return differences between the Overall Score deciles in Fig. 1, where the top panel plots the time series of monthly alphas for Decile 1 and Decile 10 portfolios, and the bottom panel displays the time series of average Overall SR Score of the two portfolios.

Both Fig. 1 and Table 3 show consistent and statistically significant differences 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 0.32% lower per month than that for 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 SR Score (0.15% per month). This is consistent with Hypotheses 1 and 4 that either preference shocks or stochastic shocks to firms' production lead to higher alphas for high-SR firms. Anecdotally, the three individual SR categories that exhibit statistically significant alpha differences also constitute the bulk of the investing criteria used by professional money managers. According to the aforementioned USSIF report, four out of seven top SRI considerations for institutional investors are related to Environment, Governance, and Product categories.[8]

With regard to the time series, Fig. 1 shows that the alpha difference between high and low Overall SR Score firms is consistent over time.[9] Except for a brief period between 2008 and 2010, the alphas for the Decile 10 portfolio is consistently higher than that of the Decile 1 portfolio. The bottom panel also demonstrates a similar story for the actual SR scores: the Decile 1 average Overall Score remains consistently negative and is either unchanged or very slowly drifting downwards, while the Decile 10 Overall Score seems to have a very slow upward trend over time. These findings lead us to reject Hypotheses 2 and 3 in the previous section.

[8]USSIF reports "Top 10 ESG Considerations for Institutional Investors" in 2012. Out of the top 10 criteria, three are not covered by our SR categories (Iran, Northern Ireland and other terrorist regimes). Among the seven remaining criteria, three are related to Governance (executive pay , board issues and political contributions), and one is related to Environment and Product (climate change and carbon footprint). The remaining criteria (Sudan) is related to Human Rights, our result for which is not statistically significant.

[9]The same pattern can be seem for the other three categories as well as the Human Right category.

3.3. These Alphas Are Temporary

Next, we examine the source of alpha difference between high- and low-SR firms, disentangling Hypotheses 1 and 4. Hypothesis 1 ascribes the alpha difference to temporary preference shocks, whereas Hypothesis 4 attributes this to risk premia for SR-related risks not measured by the researcher. Under Hypothesis 1, in order for the overall alpha differences to persist, new SRI "hot spots" must be discovered over time such that as the alpha difference in one SRI area dissipates, the preference shocks are transferred to a new area, leading to continued alpha differences. In Fig. 3, we plot the timeseries of the long-short (high SR-low SR) portfolio alphas in the three SR categories with statistically significant alpha differences (from Panel A of Table 3): Environment, Governance, and Product. Here, we see a clear transition between SR categories. Prior to 2003, both Environment and Product categories exhibit significantly positive alphas. However, those alphas quickly dissipate afterwards. Interestingly, precisely around the same time, the Governance category, which has zero alphas before, starts to exhibit significant positive alphas. Anecdotally, this transition corresponds to the passage of the Sarbanes-Oxley Act in 2002, which has put renewed public focus on corporate governance and disclosure, after a series of high-profile scandals. Our results therefore suggests that overall positive the SRI alphas has shifted from environment/product to corporate governance around 2003.

3.3.1. Relation to Previous Results

In contrast to our results, findings in both Hong and Kacperczyk (2009) and Geczy et al. (2005) provide partial support for Hypothesis 2. First, Hong and Kacperczyk (2009) construct a static portfolio of firms with substantial business interests in alcohol, tobacco and gambling industries. They find that such firms, which overlap our bottom decile firms in the Sin category, 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.[10]

In addition, Geczy et al. (2005) utilize a specific belief structure where investors have fixed prior beliefs about fund managers' skills and about 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.[11] 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 costly endeavor, our results in Table 3 indicate that this might not necessarily be the case at the individual firm level. To reconcile the two sets of results, we conduct two tests. We first obtain from CRSP quar-

[10] 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.

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

terly 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.[12] 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 by the ESGSTAT database. We then plot the time series of mean Overall SR Score for the GSL and control samples in Fig. 2.[13] This figure shows that the Overall SR Score is very similar between the GSL and control samples in the 2002-2011 period. The finding suggests that, 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, to assess the average performance of SRI and other mutual funds, we form equally-weighted portfolios of the SRI funds reported by Geczy et al. (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 4. We find that, even for the original GSL sample period, 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 our definition, the self reported SRI mutual funds are essentially the same as other funds in their investment objectives and realized performance in a model-free setting, and there is no discernable performance differences to support the alternative hypothesis of (costly) SRI initiatives.

[Insert Figure 2 and Table 4 here]

[12] 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.

[13] In untabulated results, 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. Detailed results are available upon request.

3.4. Temporary Abnormal Returns Following SR Rating Changes

To further isolate the effect of investors' SRI preferences from other risk-related confounding factors, we analyze the return changes around SR Score changes at the indidivual stock level in an event study setting. To do so, for each stock and SR Category (including the Overall Score), we first identify all the years where the SR Score decrease by one point or more (i.e. lose at least one strength criteria or add at least one concerns). Because the ESGSTAT database is updated at the end of each year and covers all developments in the current year, if there is a change in the SR Score in the current year, we choose December as the event month and the whole year as the event year. Next, for all event months, we compute the monthly four-factor alphas per Eq. 3, using return and factor data in the previous 36 months. Next, for each stock in the event sample, given the Score category c, event month m in event year t, we compute the annual cumulative abnormal returns (CARs) as the sum of monthly alphas:

$$CAR_{c,t} = \sum_{j=0}^{11} \alpha_{c,m-i} \tag{4}$$

In order to minimize any potential look-ahead bias in the CAR estimation, we also compute the average difference in raw realized returns in a given year by substrating the average returns of firms that do not experience any SR Score decrease event from those of the event firms.[14] We plot the time series of the average CAR for the Overall SR Score change event from $t-1$ to $t+1$ in Fig 4 and report the average CAR for all categories from $t-1$ to $t+5$ in Panel A of Table 5. We also report the average raw return difference for the Overall SR Score change event in Panel B of Table 5.

[Insert Figure 4 and Table 5 here]

Both Fig 4 and Table 5 provide further support for Hypothesis 1. First, in the short term,

[14] We report the difference for the Overall SR Score only. Results for other categories are similar to the alphas and are available upon request.

decreases in the SR Scores are associated with significantly negative CARs persisting up to two years after the events. For example, firms with declines in the Overall SR Score have, on average, -2.19% CAR in the event year t, -3.21% in the following year $t+1$, and -2.61% the year after that $(t+2)$. The average CARs are significantly negative in the event year t for five out of the eight individual categories. The average CARs are significantly negative for six individual categories in $t+1$ and three categories in $t+2$. Moreover, none of the CARs are statistically different from zero in the year $t-1$ prior to the event year, indicating that the negative CARs in year t are indeed associated with the events. Next, in the long term, the CARs invariably revert to zero after two to three years. With the exception of the Employment category, none of the Score categories have CARs statistically different from zero beyond $t+2$. This suggests that the SRI-related preference shocks are probably temporary and is often triggered by key SR events themselves. Moreover, the raw returns in Panel B of Table 5 follow the same pattern as the CARs in the first row of Panel A, suggesting that our estimation of CARs is not subject to much look-ahead bias.

4. Economic Mechanism: Wealth-Dependent Preference Shifts

The results in the previous section establish two facts. First, high-SR stocks have higher alphas, but these alphas are temporary, varying both across time and across different SR categories. Second, high- and low-SR firms have similar cash flow characteristics. Taken together, they suggest that the alpha differences likely result 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 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 consequently afford to be SRI-conscious. That is, they can afford to temporar-

ily 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, ceteris paribus, results in higher realized alphas. 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 "good" stocks to drop, leading to lower realized alphas. We summarize this argument in the following hypothesis:

Hypothesis 5 *In the presence of unexpected preference shocks that favor high-SR firms during good economic periods of high household wealth, the alphas of such firms will be higher than those of low-SR firms during good times due to higher demand. By contrast, during bad economic times, high-SR firms will have lower alphas than low-SR firms due to decreased investor demand.*

To test this hypothesis, we construct two indicators of good economic times using cyclicality-adjusted real P/E 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 then construct three equal-weighted portfolios:

- Good: Stocks with Overall SR Rating in the top decile
- Bad: Stocks with Overall SR Rating in the top decile
- Regular: Stocks with Overall SR Rating *not* in the top decile

We then obtain the time-series of Fama-French three-factor+momentum alphas for each of the portfolio using 3-year rolling regressions on monthly returns similar to Section 3.2. We first overlay the alphas good, bad and (good-bad) portfolios with the indicators of good times in Figure 5. We report the average alphas in these sub-periods in Table 6. This figure clearly demonstrates that the alpha difference is more pronounced during good economic times (the shaded regions). 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 Figure 5 and Table 6 here]

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

$$\alpha_{p,t} = a + bI_t + \epsilon_t \tag{5}$$

where $\alpha_{p,t}$ is the (good-bad) or (good-regular) portfolio alphas. We fit the regression using monthly data between 1993-2013 and compute Newey-West standard errors by treating the periods as continuous and use 12 lags. The results are reported in Panel C of Table 6. Again, the tables confirm the observation from the figures: the coefficient estimate for I_t is significantly positive in all six specifications, indicating that demand for high-SR stocks is indeed higher during good times, as reflected by the higher alphas for the good portfolio compared to both bad and regular portfolios.

We then directly compare SRI with luxury goods. To do so, we first construct two series of luxury good consumption and luxury retail sales:

- Luxury PCE: NIPA-based (annual series) 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 averages alphas of the (good-bad) and (good-regular) portfolios and plot them alongside the luxury consumption/sales series, in Figure 6.

[Insert Figure 6 here]

From the figures above, it is apparent that SR alphas have high correlations with both luxury consumption and sales growth. First, the (good-bad) alpha 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 "good" and "bad" stocks are attributable to the demand side, i.e. to the shifting of investors' SRI preferences, rather than to cash-flow-related explanations. The luxury-good-like performance of SRI suggests that such preference shifts are related to aggregate wealth.

5. Conclusion

We provide a comprehensive analysis of the returns and risks of socially responsible investing, utilizing a unique, firm-level database on social responsibility ratings. We advance four hypothesis governing the behavior of realized and expected returns of stocks with different levels of social responsibility, and perform empirical tests that supports the hypothesis that alpha differences are induced by shocks to investors' preferences and firms' production. In an event study setting, we show that reductions in firms' social responsibility ratings lead to significantly lower annual cumulative abnormal returns of up to two years. This result, coupled with our finding that the focus of SRI evolves over time from environment to corporate governance categories, suggests further that the alpha differences are driven by temporary preference shocks that transitions between different SR categories. We provide further evidence indicating that these differences are induced by time-varying, wealth-dependent shocks to investors' preferences, which result in high-SR stocks behaving in a fashion akin to luxury goods. The alpha difference between high- and low-SR stocks are significantly more pronounced during good economic times, and is significantly correlated with both luxury consumption from NIPA and the sales growth of luxury-good retailers.

We plan to incorporate two significant additions in the upcoming version of the paper.

First, we develop a model with wealth-dependent investor preferences and take the model directly to the firm-level data to further illustrate the economic mechanism behind the SRI-related preference shifts over time. Second, to explain the shifts in SRI alphas over different categories, we plan to directly quantify firm-level SRI shocks using textual analysis methods on firm disclosures.

References

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

Cheng, I., Hong, H., Shue, K., 2014. Do Managers Do Good with Other Peoples' Money? Working Paper, Dartmouth College.

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

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

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

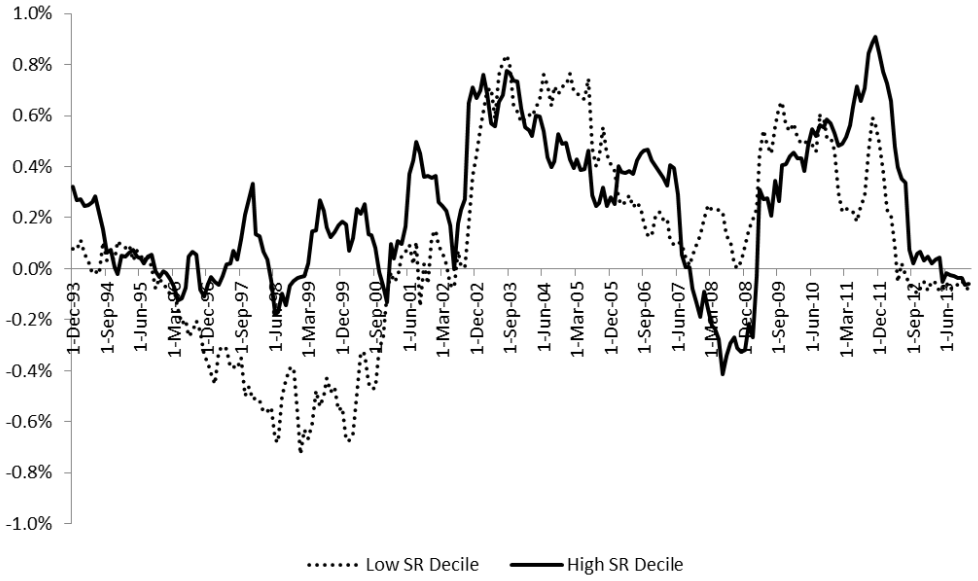
Kempf, A., Osthoff, P., 2007. The effect of socially responsible investing on portfolio performance. European Financial Management 13, 908–922.

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

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

Figure 1: **Mean Portfolio SR Scores for Top and Bottom Deciles**

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 time series of mean annual SR scores for the top and bottom deciles from 1991 to 2011. 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 1991 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.



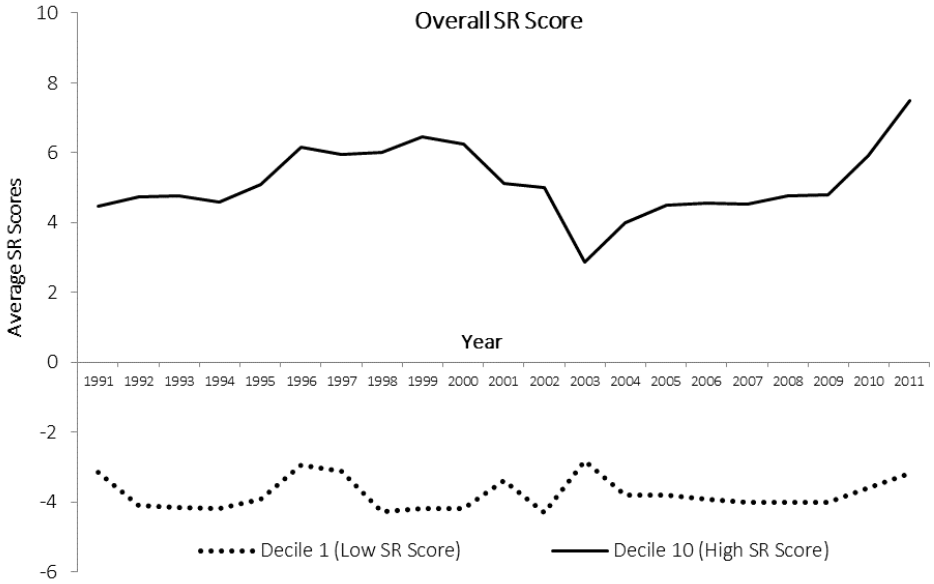


Figure 2: **SR Score for GSL Fund Holdings**

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.

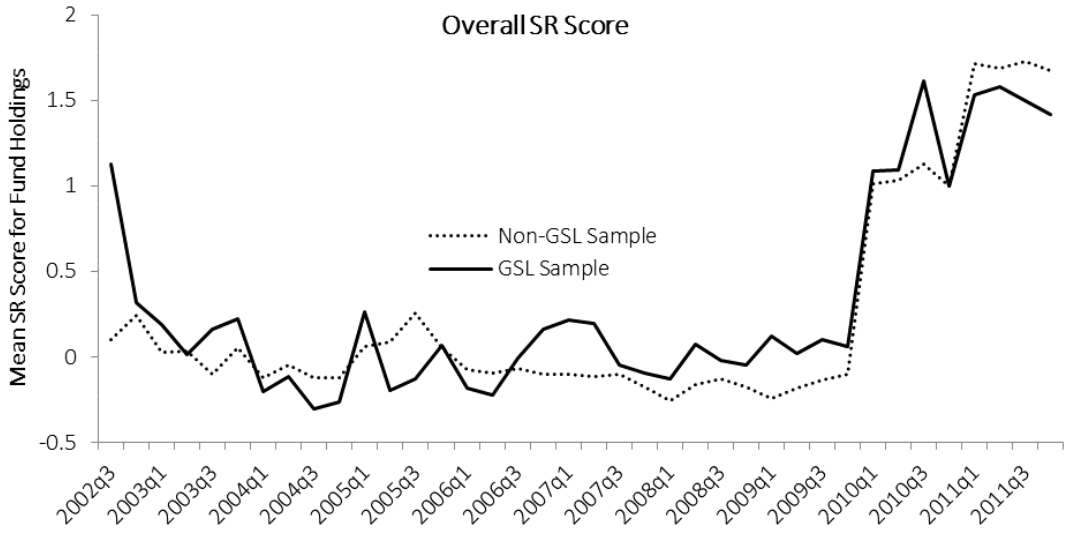


Figure 3: **Time Series of Long-Short Alphas in Different SR Categories**

The top figure displays the time series of monthly Fama-French four-factor alphas for the long-short portfolios that buy the equally-weighted Decile 10 portfolio, and sell the equally-weighted Decile 1 portfolio, in the individual SR categories of Environment, Governance and Product, from December 1993 to December 2011. 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 1991 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, where each stock is required to have at least 6 months of return data available.

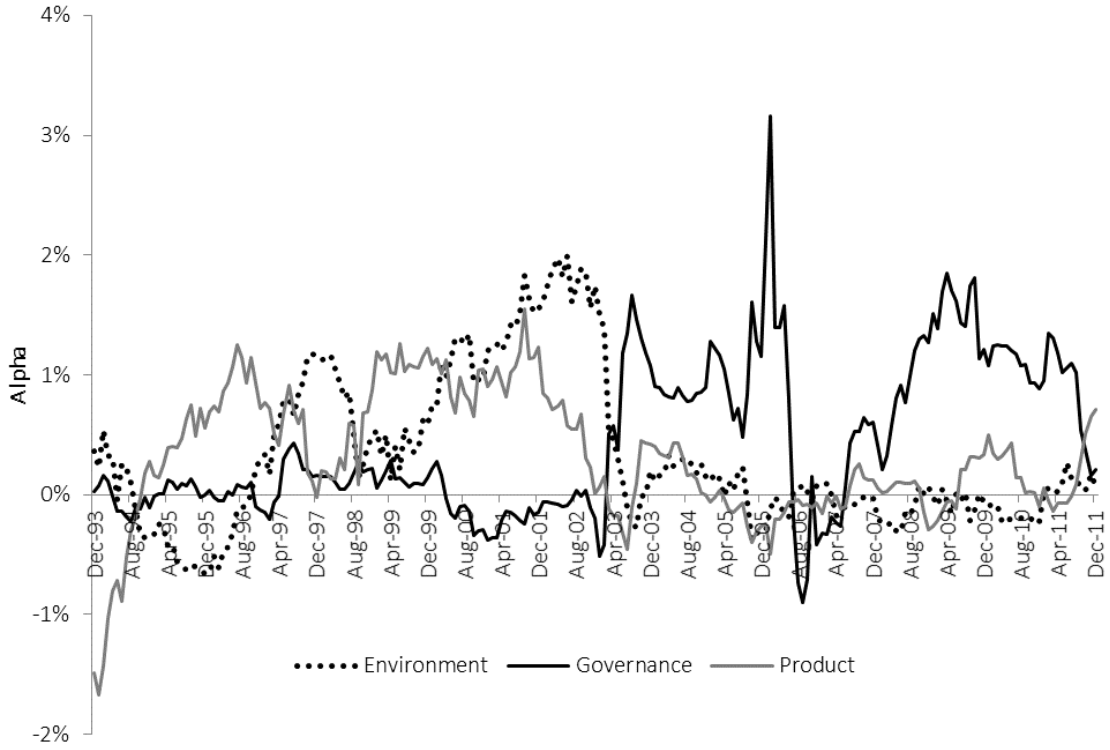


Figure 4: **Time Series of Alphas Before and After SR Score Decrease**

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 1991 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.

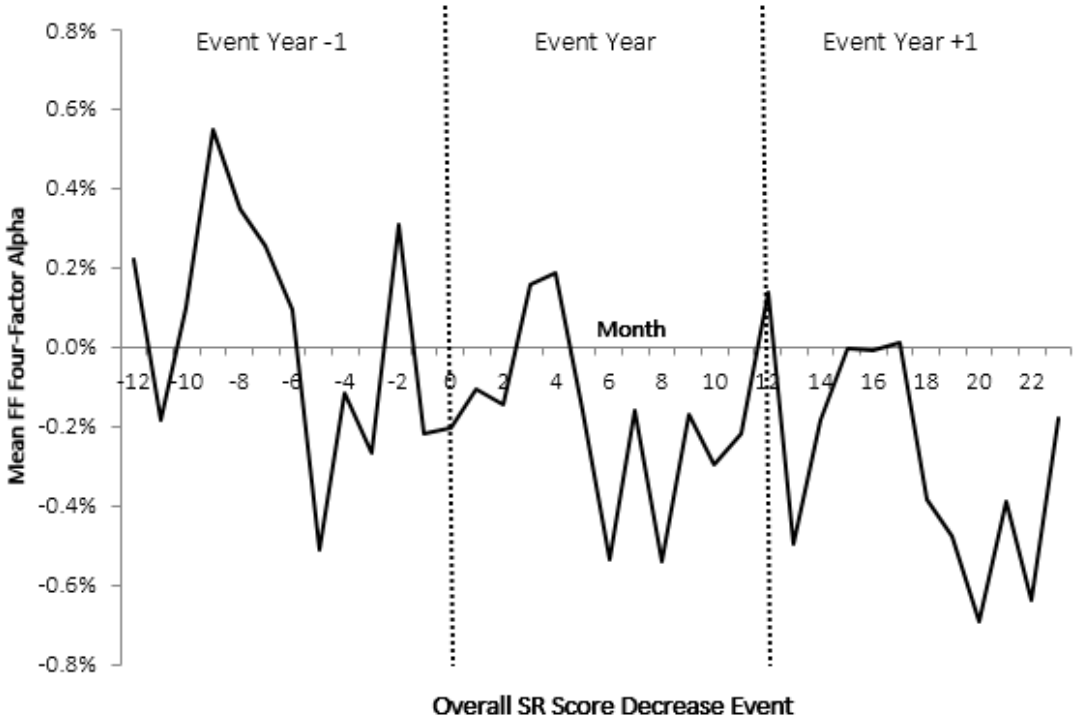
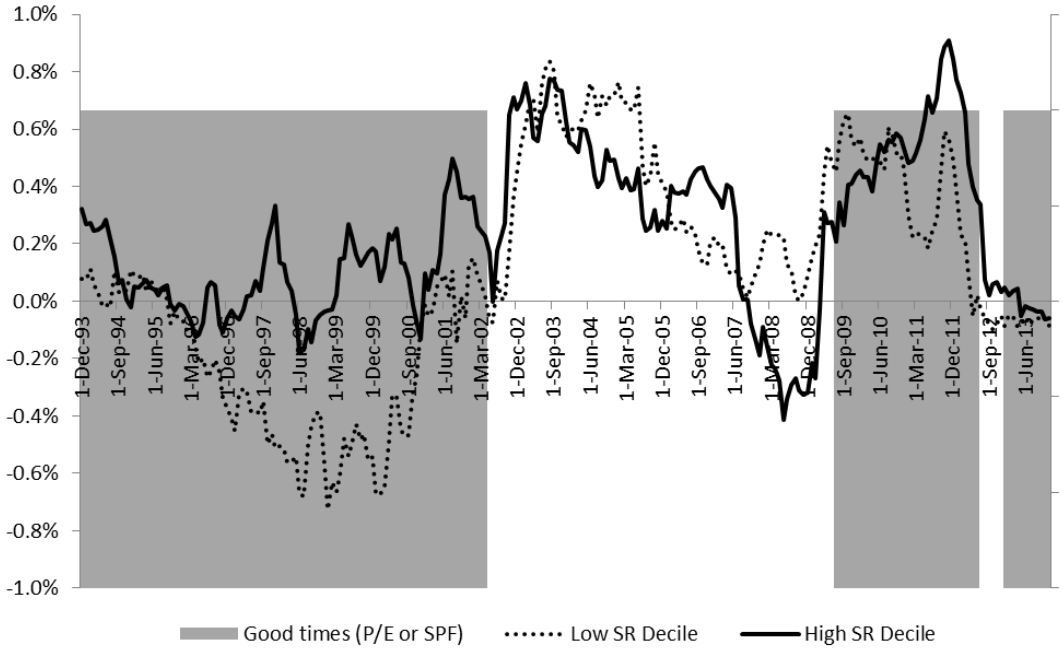


Figure 5: **Alphas for Top and Bottom SR Decile Portfolios and the (Top-Bottom) Portfolio**

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 1991 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 4 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.



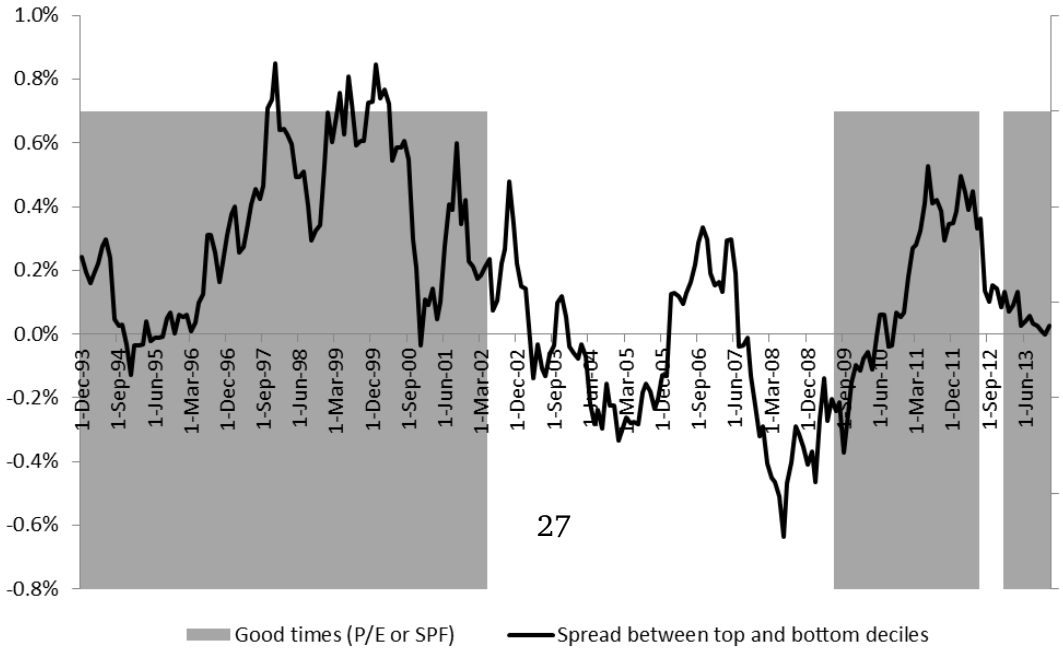


Figure 6: **Correlation between (Good-Bad) Portfolio Alphas and Luxury Consumption**

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 1991 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.

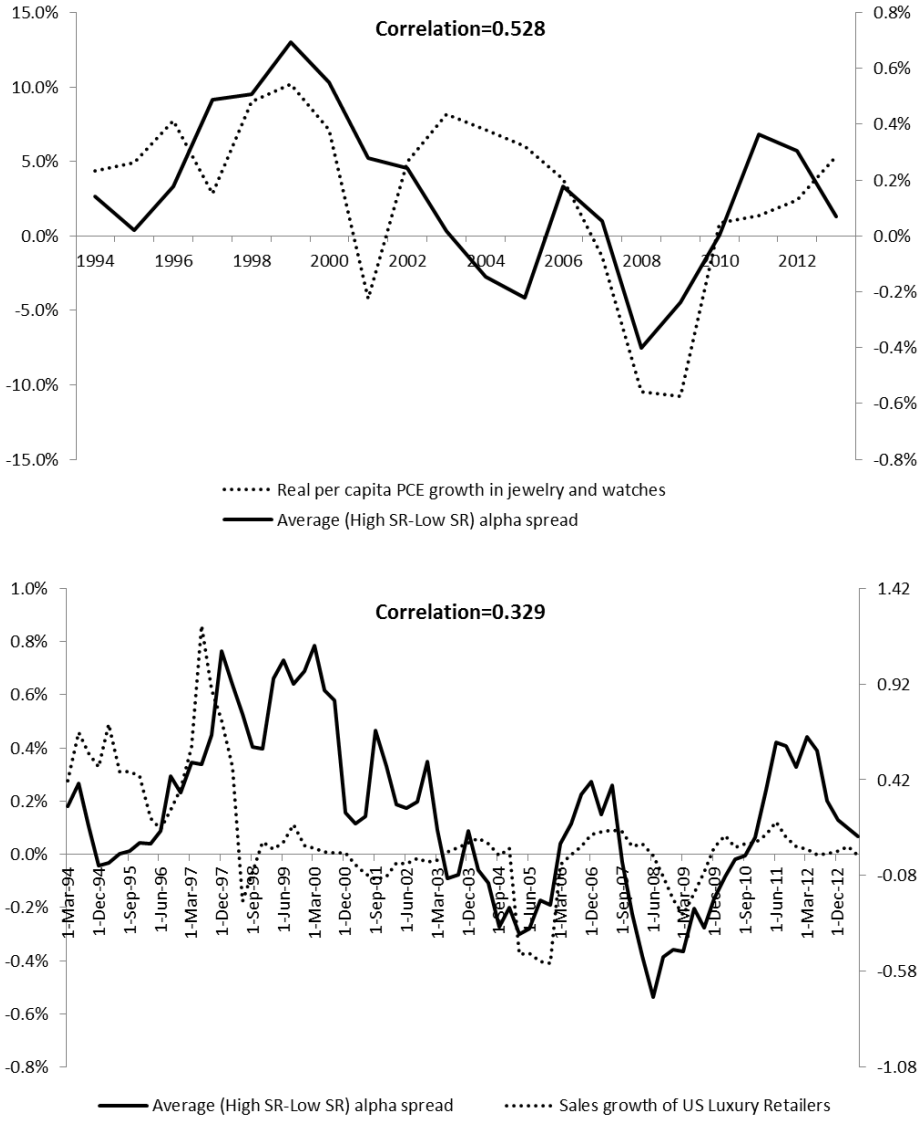


Table 1: **Top and Bottom Firms in SR Scores and Persistence in Scores**

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 1991 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 2 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 1991 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2011.

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 1991 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2011. 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: **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

Table 5: **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 point 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 point, and that of all other firms, during the same six-year span. The sample consists of S&P 500 component stocks from 1991 to 2000, Russell 1000 stocks from 2001 to 2002, and Russell 3000 stocks from 2003 to 2011. 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 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.

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 6: **Cumulative Abnormal Returns Before and After SR Score Decreases**

Panels A and B of this table presents 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 4 for details. Panel C reports coefficient estimates from Regression 5 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 1991 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 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.

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
Panel C: Alpha Regressions with Good Time Dummy						
		High SR vs. Low SR			High SR vs. Regular SR	
	(1)	(2)	(3)	(4)	(5)	(6)
Shiller	0.00318**		0.00299**	0.00239**		0.00286***
	(3.05)		(3.10)	(2.96)		(3.62)
SPF		0.00186	0.00140		0.00236**	0.00176
		(1.86)	(1.70)		(3.21)	(1.81)
No. Obs	241	241	241	241	241	241