COMPARING THE EFFECTS OF INVESTING IN PORTFOLIOS OF SOCIALLY RESPONSIBLE COMPANIES ON ABNORMAL RETURNS

BY

Spencer Koo

Submitted to Princeton University
Department of Economics
In Partial Fulfillment of the Requirements for the A.B. Degree

April 13, 2016

Abstract

In this paper, I create and analyze portfolios of socially responsible (SR) companies to see if socially responsible investing (SRI) leads to abnormal returns. I use company ratings from KLD Research & Analytics to rank companies over multiple categories and a combination of those categories. Using the simple strategies of going long on high-ranking companies and shorting low-ranking ones, I aim to: 1) discover if there exist any general return trends of SR equity investing, whether positive, negative, or neutral, 2) find out if any one particular SR trait or equity investing method that leads to particularly abnormal returns, and 3) discuss the shortfalls of some of the KLD data set. My analysis concludes that a long-short strategy over a combination of SR categories does create statistically significant abnormally high monthly returns up to 1.43%. However, the limited index range and lack of categorical rankings from KLD ultimately prevent any generalized conclusions concerning SRI, except within the S&P 500 and Russell 3000.

Table of Contents

I. Introduction 4
II. Literature Review 5
III. Data 7
IV. Methodology 9
  a. Non-Industry Specific Portfolios 10
  b. Best-In-Class Portfolios 11
  c. Various Cut-Off Portfolios 12
  d. Equally-Weighted Portfolios 12
  e. Portfolios Over Multiple Periods 13
  f. Market Model 13
V. Results 15
  a. Non-Industry Specific Portfolios 18
  b. Best-In-Class Portfolios 20
  c. Various Cut-Off Portfolios 20
  d. Equally-Weighted Portfolios 21
  e. Portfolios Over Multiple Periods 22
VI. Conclusion 22
VII. Acknowledgements 23
VIII. Exhibits 23
  a. Table I 25
  b. Table II 26
  c. Table III 27
  d. Table IV 28
  e. Table V 29
  f. Table VI 29
IX. References 30

I. Introduction

The core idea behind socially responsible investing (SRI) has existed since biblical times and has been used by groups in the United States for hundreds of years (Schueth, 2003). However, in today's the fast-paced modern world, SRI has grown immensely and represents a fair chunk of the international investing market. As of 2010, there was an estimated $3.07 trillion invested in the United States alone, nearly 21% of GDP, and an estimated $9.94 trillion invested worldwide, a majority of it in Europe (Eurosif, 2010). These investments are mainly allocated into bonds and managed monetary funds, while the SRI equity market in Europe shrank around seven percent during The Great Recession prior to 2010 (Eurosif, 2010).

As an in vogue investment sector, the definition of a sociably responsible (SR) firm continuously changes as the investor base grows. As the market continues to grow and remain profitable, the door opens to investors who may not share the traditional values behind SRI. At the moment, it is unclear if a changing investor demographic will have an adverse effect on the practice's core values or returns, but the market is gaining mainstream acceptance with little resistance (Child, 2015). Even now, it is clear that pure ethics and morality do not drive such investment, but I do not think the influx in new investors will change SRI. In fact, a significant amount of money flowing into SRI comes from large institutions with reputations to uphold such as universities and religious institutions. Not to mention, as more individual investors distance themselves from the, at times, unethical nature of finance and Wall Street, SRI offers an attractive and growing alternative for those who want to have their money work towards larger social goals, such as environmental sustainability and human rights. Hence, SRI is also known, somewhat derisively as "feel good" investing more than anything else.

This begs the question: how can one determine whether a company is considered SR without making a poor investment? Additionally, which social sectors, such as climate change, corporate governance, or employee treatment, actually make a difference in overall stock returns, if any?

II. Literature Review

Much of the literature focuses on strategies to discern if a firm is SR and the motivations behind them. The most basic screening technique, which has been seen and used by people for centuries, is the negative screen. The negative screen removes firms that are involved in areas contrary to one's beliefs, and the investments are chosen from the remaining firms. For example, in the 1980's both individuals and institutions flocked towards investment that economically pressured the apartheid system in South Africa (Schueth, 2003). While this technique can give statistically significant abnormal returns, as seen in column 1 of Table II, it surprisingly may not be the strongest way to force a company to change. Such screens can lead to unintended effects, as articulated by Gary Becker in *The Economics of Discrimination* (1957). Becker's theory states that employers with discriminatory views ultimately pay the economic price for refusing to work with certain groups of individuals (Becker, 1957).

Hong and Kacperczyk applied Becker's theory to the commonly avoided "sin stocks," which they define as those involved in alcohol, tobacco, and gambling, and they found that such stocks were often underpriced and one could obtain statistically significant returns by investing in them (Hong & Kacperczyk, 2009). Based on their results, it appears that the "discriminating" negative screen, commonly seen in SRI, does not impact "sin" firms as much as SR motivated investors might have hoped. Astutely stated by the Hong and Kacperczyk, "is socially responsible investing simply about feeling good, with little consequences for real investments?" (Hong & Kacperczyk, 2009).

Clearly, the negative screen cannot be the singular deciding factor from both a social and investment standpoint. Contrary to the exclusionary negative screening approach, positive screening creates portfolios that include companies if they behave in a SR manner. Some believe this to be a more effective screening process, especially if the SRI movement continues to grow, as it could incentivize companies to be more SR (von Wallis & Klein, 2014). Not only may positive screens be more effective, but they also lead to much greater portfolio diversity over negative screens. Thus, it is likely that these positive screened portfolios outperform the negatively screened ones (Revelli & Viviani, 2015).

Concerning the overall performance analysis of SRI, most of the literature refers to the mutual fund market, where a large portion of SRI takes place. However, as pointed out by Geczy, Stambaugh, and Levin, generalizing the economic and social effects of these funds is difficult. Mainly, this occurs because investors have inherently different market beliefs and investing philosophies. With respect to mutual funds, the major differences concern asset pricing models and managerial stock picking skill (Geczy, Stambaugh, & Levin, 2005). Geczy et al show that those who believe in efficient markets and the capital asset pricing model (CAPM) rather than managerial skill face little cost to investing in a SR fund. The restraints of SRI become costlier to the investors who believe in pricing models that take into account market loading factors such as market capitalization, the book-to-market ratio, and momentum factors. The SR fund is the costliest to those who strongly believe in managerial skill (Geczy, Stambaugh, & Levin, 2005). Therefore, though there is a substantial amount of literature on the subject, it does little to answer the proposed question concerning generalized performance for SR equity portfolios.

One article by Kempf and Osthoff, which this paper is based on, specifically addresses that vital question. They find statistically significant abnormal returns that reach as high as 8.7% annually when screening by industry for a best-in-class portfolio, using a combination of multiple screens, and restricting the portfolios to the highest ranking SR companies (Kempf & Osthoff, 2007). This is the approach that I take in this paper, from which, I ultimately come to very different conclusions. Many of the differences are due to shortcomings in the data itself, which is a fact that Kempf and Osthoff fail to recognize and delve into in their work.[1]

III. Data

In order to screen for various portfolios, I use the social environmental, social, or governance (ESG) company ratings compiled by KLD Research & Analytics (KLD)[2] from 1991 to 2003 (MSCI, Inc., n.d.). In August of each year, they took the companies from the S&P 500 and DS 400, averaging 669 stocks annually[3], and rated them on seven SR categories: environment, community, human rights, employee relations, diversity, product, and corporate governance. All of the categories have subcategories for both strengths and controversies with each ranked in a simple binary system. A company gets a one for any subcategory if it has that

[1] I use a similar methodology based on that used by Kempf and Osthoff. In their original work, they did not articulate the specifics of their methodology that can change the results drastically, as seen in my work.
[2] KLD Research & Analytics is owned by MSCI, Inc. as of 2010. I cover the years from 1991 to 2003 so I largely refer to the KLD methodology in collecting the data (Demos, 2010).
[3] In 2001 and 2003, KLD expanded their ratings to more both the Russell 300 and the Russell 1000 indices; however, for consistency within my study, I use only the companies in the S&P 500 and the DS 400 through 2003. This proved to be one of the more challenging aspects of the study due to KLD's poor organization and presentation of its data. The only way to get specific KLD data is by using CUSIPS or tickers, both of which are either non-uniform or can be reused) so I proceeded by obtaining entire index lists for the S&P 500 (Hines & Bordelon, 2016) and DS 400 (Princeton Data and Statistical Services, n.d.) and carefully filtering dates and running the tickers through the KLD database.

trait and a zero if it does not (MSCI, Inc., 2015). This rating system disregards whether the trait is a strength or controversy.

The total number of subcategories changes every year as various traits are added, removed, consolidated, or broken up into more specific subcategories. Generally, each category's meaning remains consistent with most investors' definitions with the exception of corporate governance. According to Gompers, Ishii, and Metrick, corporate governance defines the "power-sharing relationship between investors and managers" (Gompers, Ishii, & Metrick, 2003). However, the KLD corporate governance definition concerns issues such as leadership compensation and bonuses, involvement in government affairs, and public support for SR government policies. Also encapsulated in the definition is whether a company owns as little as 20% of a SR company (as defined by KLD ratings), accurately reports its sustainability efforts, and has "a unique and positive corporate culture" (MSCI, Inc., 2015). Because of the lack of conformity to the typical definition of corporate governance, I do not include this category in my analysis.

KLD also investigates whether companies are involved in the so-called "sin" industries as mentioned by Hong and Kacperczyk. They extend the category to include companies involved in any industry concerning alcohol, gambling, military, nuclear, tobacco, and firearms[4] (MSCI, Inc., 2015). These negative, or exclusionary, subcategories follow a simple binary system as well. If a company is involved in any of these areas, it receives a one and vice versa.

The financial data used to obtain the overall portfolio returns comes from the Center for Research in Security Prices (CRSP) (CRSP, n.d.). I use the monthly holding period returns including dividends, monthly price, outstanding share data, and the standard industrial

[4] The firearms subcategory was included from 1999 forward.

classification (SIC) codes to create the best-in-class portfolios, as described next in the methodology section (CRSP, n.d.).[5]

The data for the market models and the 10 industry code ranges come from Ken French of Dartmouth University (French, Description of Fama/French Factors, 2016). I use the risk-free rate (R_f), [6] his CRSP market portfolio performance measure (R_m-R_f) as a benchmark for comparison, and the factor-loading variables: SMB (market capitalization ratio), HML (book-to-market ratio), and MOM (momentum) to run the Carhart four-factor model (French, Detail for Monthly Momentum Factor (Mom), 2016). The industry code ranges are used to create the best-in-class portfolios (French, Detail for 10 Industry Portfolios, 2015).

IV. Methodology[7]

In each of the next subsections, I will present detailed explanations of how I use the data from the last section and apply it to create the various portfolios. Unless explicitly stated otherwise, each subsection builds off the other. In other words, if I explain how the negative screen is performed, then it will not be described again in further subsections unless applied differently.

The screens are applied annually to all the companies in the S&P 500 and the DS 400. After the screens select the portfolio's companies, each company's monthly returns are value-weighted based on its monthly market cap relative to the portfolio.[8] The weighted sum of all the

[5] Obtaining all the correct data from CRSP was not an easy task as well. CRSP returns data for each month specified within the time range and tickers can be reused so I had to match up the tickers from KLD with the PERMNO identifiers from CRSP. In the event of companies with multiple PERMNOs, signifying two classes of stocks, then both were kept for the screening process.

[6] I obtained the risk free rate from Ken French's website (French, U.S. Research Returns Data (Downloadable Files), 2016).

[7] The overall methodology is taken from Kempf and Osthoff with alterations in details (Kempf & Osthoff, 2007).

[8] Because KLD publishes new ratings each year, these scores are used to create portfolios for the following year. Thus, my portfolios' returns cover 1992 to 2004. All of the portfolios are updated annually and do not change during the year. If a company goes out of business or is taken over and ceases to exist in the CRSP database, then this is accounted for through value-weighting.

companies' returns is the total monthly return for the portfolio. These weighted monthly returns combine to form all of the total monthly portfolio returns over the course of the 13-year period. In most cases, three portfolios are made per screen: high-ranking, low-ranking, and long-short, which goes long on the high-ranking portfolio and shorts the low-ranking portfolio.[9]

A. Non-Industry Specific Portfolios

For non-industry specific portfolios, I run negative, positive, and combination screens to make value-weighted high-ranking, low-ranking, and long-short portfolios from 1992 to 2004.

For the negative screen, the high-ranking portfolio is created by removing any company involved in any of the controversial industries. The low-ranking portfolio is formed by only including companies that are involved in at least one of the controversial areas.

For the positive screens, portfolios are made for each of the categories individually. In order to do this, the binary complement is taken for each of the controversy subcategories;[10] in other words, all of the subcategories are now considered strengths. This is done for all of the positive and combination screens moving forward. Within each of the six positive categories each company's strengths are summed and divided that category's total possible subcategories. This value allows the companies to be ranked in terms of social responsibility. The high-ranking portfolio for each category is obtained by keeping all those companies at or above the 90th percentile. Conversely, the low-ranking portfolio for each category is composed of those companies at or below the 10th percentile. [11] Each companies' ratio of strengths within each

[9] For the long-short portfolios, the high-ranking portfolio becomes non-inclusive while the low-ranking portfolio remains inclusive. This is to prevent investment overlap. The reasoning for including this in all of the percentile cut-offs rather than only the 50th percentile will be made clear in the results section.
[10] This does not include the negative screens. Recall that the positive subcategories contain strengths and controversies and are scored using a binary system irrespective of whether they are strengths or controversies.
[11] In Stata, I used the upper and lower bounds of the percentiles for the high- and low-ranking screens respectively. The choice had no effect on the results because the upper and lower bounds were always exactly at the percentile mark.

category is recalculated each year to take into account the frequent changes to the number of subcategories.

The combination one screen divides the sum of all the available subcategory strengths over the total number of subcategories possible and similarly ranks the companies highest to lowest. The high-ranking and low-ranking portfolios are created at the 90^{th} and 10^{th} percentiles respectively. Combination two first applies the negative screen and then applies the combination one screen on the remaining companies. The high-ranking and low-ranking portfolios are again created using the 90^{th} and 10^{th} percentiles respectively.

B. Best-In-Class Portfolios

Compiling the best-in-class portfolios involves one more step than the non-industry specific ones. Using the SIC codes from CRSP and the industry ranges from Ken French, the companies for each year are separated into the 10 industries.[12] Once separate by industry, the high-ranking, low-ranking, and long-short portfolios are made using the positive and combination screens similarly to the non-industry specific portfolios. However, rather than screen all of the companies for each month together, the companies are screened by their respective industry sector.

Once screened, each of the singular industry portfolios is combined into an industry-weighted monthly portfolio. This is achieved by dividing each sector's market cap over the entire month's market cap. Then, this industry specific ratio is applied equally to each firm within that respective sector. The resulting sum of the industry-weighted returns gives the overall monthly

[12] The industries are: consumer non-durables, consumer durables, manufacturing, energy, high technology business equipment, telecom, shops/shopping, health, utilities, and other, which includes mines, construction, building management, transportation, hotels, bus services, entertainment, and finance.

portfolio return, which is subsequently value-weighted in the way described above in subsection A of methodology.

C. Various Cut-Off Portfolios

The portfolios at various cut-offs use both the non-industry specific method as well as the best-in-class strategy. The key difference here is that the categorically ranked companies from the positive and combination screens are cut-off at different percentiles. The new cut-offs are run at the 5^{th} (95^{th}), 25^{th} (90^{th}), and 50^{th} percentiles for the low- and high-ranking portfolios respectively. Only the long-short portfolios are created for this subsection, but it is comprised of the other two portfolios. Note that the negative screen is not run as it uses an exclusionary screen rather than a percentile cut-off. Keep in mind, as mentioned in footnote 9, for the long-short portfolios, the high-ranking portfolio becomes non-inclusive at the percentile cut-off while the low-ranking portfolio remains inclusive to avoid investment overlap.

D. Equally-Weighted Portfolios

The equal-weighting is applied instead of value-weighting by dividing one by the total monthly cap and applying those equal weights to each company creating the total monthly portfolio return. Parallel to the value-weighting, each month is weighted in similar fashion. The purpose, as stated by Kempf and Osthoff is to see "whether our results are sensitive to [the] portfolio weighting scheme" (Kempf & Osthoff, 2007).

It should be noted that equal-weighting does not take the place of industry-weighting for the best-in-class portfolios. In some years, certain industries have single digit numbers while others have upwards of 170. Without testing equal-weighted industries, it is clear that this would skew the data significantly if the lesser represented industry sector held the same weight as the other. Remember, I am looking for abnormal returns due to SRI, not because a particular industry had consistently high returns and more capital invested in it.

Long-short portfolios are made for the negative, positive, and combination screens using the non-industry specified method. For the best-in-class method, long-short portfolios are formed for the positive and combination screens are created.

E. Portfolios Over Multiple Periods

The last set of portfolios cuts the period into two almost equal sub-periods: 1992 to 1997 and 1998 to 2004. Again, only long-short portfolios are formed, and the negative screen is only applied to the non-industry specific portfolios. The purpose here is to see if the results hold consistently over time, or if there was one particular time period that can be attributed to any abnormal returns.

F. Market Model

With the monthly returns for each portfolio, I use the Carhart four-factor model to estimate the excess abnormal returns (alpha) relative to the benchmark CRSP market portfolio. Built off the Fama-French three-factor model, Carhart has been commonly accepted in the world of finance and offers a closer look at the types of companies each portfolio consists of. The model uses the market portfolio's return in excess of the risk free rate, market cap ratio (SMB), and book-to-market ratio (HML) (Fama & French, 1993). SMB shows whether the portfolio leans more towards large or small cap companies relative to the market portfolio. HML gives insight into whether the portfolio is composed of more value or growth stocks, again, relative to the market portfolio (French, Description of Fama/French Factors, 2016).

Bauer, Derwall, and Otten point out in their work on SRI in Canadian mutual funds that SR funds are particularly prone to the small-firm effect, generally made up of smaller cap and growth stocks, and are less exposed to the market portfolio overall (Bauer, Derwall, & Otten, 2007). While this describes funds, controlling for these factors gives a more accurate representation of any abnormal excess returns.

The Carhart model's fourth factor comes from his work studying mutual fund returns. He discovered that much of these returns can be attributed to the three-factor model plus a fourth factor, momentum (Carhart, 1997). My portfolio's composition will not drastically vary year to year so I will use Carhart's suggested four-factor model:

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{mt} - R_{ft}) + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}MOM_t + \varepsilon_{it}$$

Where $R_{it} - R_{ft}$ is the portfolio's return in excess of the risk free rate, α_i is the portfolio's abnormal return not explained by the market portfolio or any independent factor-loading variables, $R_{mt} - R_{ft}$ is the market portfolio in excess of the risk free rate, SMB_t is the market cap ratio,[13] HML_t is the book-to-market ratio,[14] MOM_t is momentum,[15] and ε_{it} is the error term.

Due to the regression using the market portfolio as a benchmark, it is highly plausible for a normal ordinary least squares (OLS) regression to autocorrelate. In other words, the regression correlates with its own variables (Buchanan, 2011). This serial correlation means that finding the error for one observation gives insight into the error for the next observation, meaning the observations are not independent (Murray, 2006). This is very important because inaccurate error terms could falsely show a variable in the regression as having a statistically significant effect when it in fact does not and vice versa. Thus, it could potential ruin any results.

[13] SMB stands for small minus big. It is calculated by taking the average return of three small cap portfolios and subtracting the average return of three large cap portfolios (French, Description of Fama/French Factors, 2016). The larger and more positive the coefficient on SMB means that the portfolio leans towards small cap firms and vice versa.

[14] HML stands for high minus low. It is calculated by taking the average return of two value portfolios and subtracting the average return of two growth portfolios (French, Description of Fama/French Factors, 2016). The larger and more positive the coefficient on HML means that the portfolio leans towards value stocks and vice versa.

[15] Momentum is calculated by taking the average return of two prior high return portfolios and subtracting the average return of two prior low return portfolios (French, Detail for Monthly Momentum Factor (Mom), 2016).

Normally using a robust standard error will correct for any heteroskedasticity (Murray, 2006). However, given that this data is a time series, it is more appropriate to use a Newey-West regression. These types of regressions "fix up the OLS standard errors so they are no longer biased but remain inefficient" (Wadsworth). This is achieved by lagging the dependent variables behind in the regression thereby furthering the distance in which the serially correlated terms can interact and rendering the effect as negligible as possible (Newey & West, 1987). Because this study uses monthly return data, I will use a lag factor of 12 for my regressions to best estimate the standard error.

V. Results

Before discussing the results of the regressions, it is important to note the reason why percentiles were used over percentages for the cut-offs. This choice uncovers an important flaw in the KLD datasets that Kempf and Osthoff did not recognize.

In short, there are simply not enough subcategories to get an accurate ranking within each category, which leads to a large amount of ties in rank.[16] Cutting-off at the percent mark rather than at the percentile arbitrarily excludes companies that have the same categorical ranking as others that are included in the portfolio. This leads to a bias in the found returns. Instead of focusing purely on the SR category to form the portfolio, cutting-off at the percentile leans the portfolio in whichever way the data is sorted. For example, if the data is sorted by largest market cap to smallest, then the returns will be biased towards larger cap firms. Or, if the data is sorted alphabetically, then the end results may suggest that abnormal returns occur , but only for firms whose names are in the beginning of the alphabet. There are many ways to sort through the data

[16] Recall that the categorical rankings are based on the sum of each company's total strengths within that category divided by the total number of possible subcategories for that category.

after categorical rank, but all of them will bias the results in some manner. Even if the sorting were randomized after sorting by rank, there are simply not enough companies on an annual basis to sufficiently create unbiased results. The real core of the issue lies with the KLD data.

In further detail, the largest subcategory count (with the mean in parentheses) for each of the six categories over the course of the study[17] is: 10 (9.46) for community, 11 (10.54) for diversity, 11 (10.08) for employee relations, 12 (11.77) for environment, 7 (4.92) for human rights, 8 (8) for product, and 58 (54.77) using all subcategories. While the individual categories may seem to have a sufficient number to properly rank all the companies, keep in mind that each year there is an average of 669 companies to screen. Given that large of a number, it is likely that they converge towards the middle, and the data shows that there are a substantial number of ties in rank amongst all the categories.

Table I shows the average annual number of companies left over after each positive high-ranking screen for all four of the various percentile cut-offs. For the best-in-class screens in columns two, four, six, and eight, the data represents the average number of companies per industry sector per year. Both include the standard deviation in parentheses. Only the high-ranking screens are performed for this exhibit because the low-ranking portfolio mirrors those results. The second value in Table I reveals the real percentage cut-off, the measurement that Kempf and Osthoff[18] use (Kempf & Osthoff, 2007). This value differs from percentile in reference to the cut-off point because the same number is used to describe the cut-off for both high- and low-ranking portfolios. In other words, for the high-ranking portfolio cut-off at 10% those above the top 10% remain irrespective of ties in rank, as explained earlier. In terms of

[17] Recall that the number of subcategories can change when the new ratings are published annually.
[18] In their paper, it is unclear what their exact methodology is for the cut-offs. They simply note that they cut-off at these percent marks without ever mentioning percentile or another more specific method. Therefore, I assume that they used percent.

percentile, the 10% high-ranking portfolio would be cut-off at the 90th percentile, inclusive of ties in my analysis. For example, in column three of the community screen, on average 134.21 companies remain per year in the portfolio after screening and that represents, on average, 20.06% of the total number of companies screened.[19]

Looking at Table I, the large number of ties is represented by the much larger than expected real cut-off at each percentile. A prime example is the human rights category, where, regardless of the percentile screen used, the actual percent cut-off is exactly the same for all of the non-industry specific screens. In the raw KLD dataset, the human rights category only has two subcategories in total from 1991 to 1994. Even though that is an extreme case, the results from Table I clearly show that all the categories, for both non-industry specific screens and best-in-class screens, simply do not have enough subcategories to appropriately rate these companies on SR behaviors. That said, the two combination values give real percent cut-offs very close or exactly at the percentile cut-off mark. Thusly, when combining all the categories, one gets a much better picture of whether a firm can be considered SR.[20]

A possible solution to this issue is to use non-inclusive percentiles for the screens. However, when I created the positive screen portfolios non-inclusive of ties at the percentile mark, there were generally as few as single digit numbers of companies for even the 50th percentile cut-off. Even though the regressions reveal some very high and significant alphas, they would not be close to representing the markets accurately and would give very little insight into SRI as a whole, which is the goal of this study.

[19] The negative screen is not included because its screen is exclusionary and does not depend on cut-offs.
[20] Clearly, the data presents a dilemma jeopardizing the results. Kempf and Osthoff do not address this issue in their work, which lends me to believe that they either did not recognize it or circumvented the issue in some manner and failed to mention it (Kempf & Osthoff, 2007).

The most feasible solution comes from the data source, KLD or MSCI. Looking at their 2014 Data Manual, the only category that has greatly expanded its number of subcategories is environmental issues (MSCI, Inc., 2015).[21] The others still contain around the same number of subcategories as they did at the end of 2003. If expanding the number of subcategories is not possible, I propose using a scoring system from one to 10 or one to 100. This would add much more variety to the average scores within each category and even more specificity for the combination screens' scores.

Keeping the stark limitations of the data in mind, in the following subsections the results of the Newey-West regressions are presented. I mention the individual category screens and various cut-offs; however, I focus mainly the negative and combination screens since they most accurately reflect screening based on SR behaviors.

A. Non-Industry Specific Screens

Starting with the negative screen, the high-rated portfolio has marginally high monthly abnormal returns (alpha) of 0.12% significant at the five percent level and relative to Ken French's benchmark market portfolio made from CRSP data (French, Changes in CRSP Data, 2016). The low-rated portfolio slightly underperforms against the CRSP benchmark. Since the negative screen contains all of the stocks used in the KLD universe and there are no drastic underperformances compared to the CRSP benchmark, focusing on the long-short portfolios will offer more insight into the potential for return within the KLD universe (Kempf & Osthoff, 2007). It is, however, necessary to keep in mind the limitations of the data. Again, for the long-short portfolios, the high-rated portfolio becomes non-inclusive and the low-rated portfolio

[21] This is most likely due to large public and investor interest in environmental protection and sustainability. According to Berry and Junkus, "For…investors, environmental and sustainability issues dominate as the major category associated with SR investing" (Berry & Junkus, 2012).

remains inclusive. This is to avoid potential overlap as seen in Table I. The issues with KLD data will affect all of the screens except for the negative screen, which does not use percentile cut-offs. With that said, the negative screen long-short portfolio gives relatively large monthly alpha of 0.17% significant at the five percent level.

Overall, the singular category positive screens do not seem to be of much use despite some very high and statistically significant alphas. However, by comparing the real cut-offs from Table I, some of the statistically significant alphas offer some insight. The high-rated community portfolio has a real percentage cut-off at around 20% for the 90th percentile, and it offers relatively large returns. That portfolio seems to be composed mainly of larger firms, which makes sense given some community programs to help families and neighborhoods of employees. This could prompt a more efficient and harder working staff.

I believe that the results of the long-short portfolios are generally anomalies given the issues with the data. However, the two combination screens show very impressive alphas both significant at the one percent level. This is further proof that ranking firms based on SR traits works over a more extensive ratings system. While the data for the combination screens turned out nicely, the same cannot be said for the singular positive screens.

As far as the factor-loading variables are concerned, the only noticeably consistent portfolio makeup across the screens in Table II is the book-to-market ratio (HML). With the exception of the community screen, the high-ranking portfolio generally has an HML coefficient that is significantly lower than the low-ranking portfolio of the same screen. This reveals that the high-rated portfolios' compositions lean more towards growth stocks rather than value stocks.[22]

[22] The difference between growth and value investing generally has to do with an investor's personal market philosophy. Investors who prefer growth stocks, also known as glamor stocks (Investopedia, n.d.), tend to believe that the current market price reflects all relevant information and is a fair price. They trust that the price includes the growing nature of the firm and that the price will invariably rise as the firm grows (Dow, 1998). Investors who

B. Best-In-Class Screens

The best-in-class screens seem to be in overall worse shape given that the data is cut up into smaller industries before being screened. From Table I, the average real percent cut-off is for each industry sector. Thus, while it may seem like a higher percentage than the non-industry specific screens, the overall number is actually smaller, which certainly is reflected in Table III.

There is, however, another great example of my argument for ways to improve the data set when looking at the two combination screens. Though the high- and low-ranking portfolios do not give positive returns, the long-short portfolios perform quite well at 0.24% and 0.46% monthly for combinations one and two respectively. Both are significant at the one percent level.

Overall, there seem to be no overarching trends on the factor-loading independent variables in this case.

C. Various Cut-Off Screens

Table IV is best compared right next to Table I. Table IV gives the long-short alphas from the regressions for the various percentile cut-offs. Overall, in the best-in-class portfolios, the only screens that give any positive statistically significant return are combinations one and two. Combination two is positive and significant at the one percent level at the 95th and 90th percentiles, showing 0.24% and 0.46% monthly excess return respectively. Combination is significant to the one percent level at the long-short 10th percentile cut-off, showing 0.24% monthly alpha.

The returns are higher for the non-industry specific portfolios. Again, it seems to be an issue of the number of subcategories available during screening. Nevertheless, the long-short portfolios for the non-industry specific portfolios offers some unexpected results. The diversity

---

prefer value stocks believe that some stocks are underpriced or undervalued. Generally those firms have “low price-to-earnings, price-to-book, and price-to-sales ratios, and high dividend yields” (Dow, 1998).

screen at the long-short 25th percentile gives quite high returns of 1.67% monthly alpha. In terms of real cut-off, Table I shows it to be just shy of 26%.

Looking at the combination screens for the non-industry specific, it seems that overall, the more selective the screening process (i.e. the lower the cutoff), the better the portfolio performs. However, looking again at Table I, it seems that there is a limit to the selectivity. Once the number of firms gets too low, the portfolio faces a steep drop off in return and significance as seen for combination one in panel A under column (1). Given the data issues, it is difficult to tell whether the particular exclusivity point that earns the most is different for every screen or more a matter of getting the right amount of firms to regress over. It is most likely both.

D. Equally-Weighted Screens

While there were some very drastic changes to the results for the 10th percentile under the equal-weighting scheme, there are some peculiar results to be in the best-in-class portfolios in it has the only result that stays exactly the same. The product screen for best-in-class remains at -0.46% significant at the one percent level regardless of the weighting scheme.

Much of these results are sensible given the number of portfolios that remain relatively static in terms of real percentage cut-off. For example, the human rights screen's alpha does not move much from prior Tables' results. When comparing to Table I, there are so few subcategories in the human rights main category that the number of firms in each portfolio remains almost entirely static regardless of the changes applied. With such a large number of upwards of 90% included in the portfolio, equally-weighting would have little to no effect.

On the other side of the spectrum, some of the very drastic changes like the non-industry specific diversity screen, which falls from 0.70% alpha to -0.86% alpha. Both are long-short portfolios at the 10th percentile and statistically significant at the one percent level. Although from table I, there are a reasonable number of firms regressed over at the 10^{th} percentile cutoff, the weighting scheme clearly has a drastic effect on some of the categories. Perhaps combining some of the largest firms in the S&P 500 and the smaller social index of the DS 400 could create such a disparity.

E. Screens Over Multiple Periods

Interestingly, the latter period from 1998 to 2004 gives much stronger significant returns over the earlier period. This perhaps lends itself to overall stronger market performance? Or just stronger performance of the KLD universe of stocks over the CRSP market universe.

Though not statistically significant, with the exception of combination one best-in-class, the two combination screens outperform most of the singular screens. This does give some credence to using a broader base to analyze the SR of firms. Nevertheless, it does seem that both the negative screen and the diversity screen under non-industry specific portfolios perform consistently well over the entire timespan, all the while statistically significant at the one percent level.

Another possibility lies in the question I proposed in the introduction. Perhaps, with the increasing popularity of SRI, more high profile and well-performing companies started being more socially responsible. The non-industry specific environment alpha went from -3.70% to 1.74% in just six years (both statistically significant at one percent). Also, the general upward trend of the combination portfolios could suggest that overall more firms in all areas are being more socially conscious; however, the massive rise in the environment alpha is the most compelling evidence for that theory as it is one of the main focuses of SR around the world and especially in public policy.

VI. Conclusion

As a whole, I do not think that these results can be generalized to SRI due to the pitfalls in the KLD data set. Nevertheless, I believe that there are some implications to the data that can be used in further studies or revisited with updated data. First and foremost, the combination screens generally outperform their positive screen counterparts by a significant margin. Secondly, using an industry mixed portfolio seems to hurt portfolio performance more than anything (another difference from the original Kempf and Osthoff study). Third, though there is literature concerned that negative screening policies lower diversity of the SR portfolio and do not discourage "sin" firms from their behavior, these results show generally consistent high and significant monthly alphas.

If I were to continue on with this study, the best course of action would be to expand the study to later years or to start over using all more indexes that KLD (and MSCI) expanded to after 2001. Given the number of studies concerning this fast-growing field, there are most likely other datasets and methods used to rank companies based on SR qualities. I think that combining one or more of these ratings systems into a more in depth and rigorous hybrid would certainly mitigate 1) the obvious issue of a lack in subcategories and a shallow binary ratings system and 2) multiple ratings would balance out the counterargument that these ratings are just one other firm's opinion and cannot possibly be used to judge other companies' SR statuses.

Overall, the field is bound to grow further given the current state of politics, domestic and foreign, as well as the ever present threat of major climate changes. With that said, more data will surely become available in the near future to return to this study and offer a more complete view on whether or not SRI does lead to abnormally high return.

VII. Acknowledgements

I would like to thank Dr. Motohiro Yogo for his guidance, assistance, and patience throughout the year. I would also like to thank Mr. Todd Hines for help obtaining and manipulating all the necessary data. Additional thanks to Mr. Alexander Kempf and Mr. Peer Osthoff for their original paper.

VIII. Exhibits

A. Table I

Average Annual Number of Companies and Average Annual Percent Real Cut-Off (including Standard Deviation) for Non-Industry Specific and Best-In-Class (BIC) Portfolios From 1992-2004

Percentile Cut-off		5%		10%		25%		50%	
		No BIC	BIC	No BIC	BIC	No BIC	BIC	No BIC	BIC
Categories		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Community	Companies	57.05	6.79	134.21	12.86	279.71	44.44	633.19	99.85
		(12.48)	(5.55)	(49.54)	(9.65)	(179.64)	(31.92)	(20.21)	(48.92)
	Percent Cut-off	8.54	10.42	20.06	12.34	41.32	41.10	94.65	93.25
		(1.89)	(10.82)	(7.40)	(8.64)	(25.83)	(28.58)	(3.04)	(10.46)
Diversity	Companies	45.19	4.97	95.25	10.84	310.73	46.46	502.57	76.71
		(11.46)	(3.83)	(29.85)	(7.78)	(179.62)	(41.97)	(137.19)	(48.16)
	Percent Cut-off	6.75	7.97	14.23	10.27	46.49	37.16	75.02	68.41
		(1.71)	(8.01)	(4.45)	(6.04)	(26.96)	(25.03)	(20.45)	(23.89)
Employee Relations	Companies	67.17	9.84	156.15	18.08	291.41	47.26	556.16	87.47
		(31.66)	(8.74)	(48.42)	(13.27)	(155.49)	(39.73)	(32.38)	(46.81)
	Percent Cut-off	10.04	13.30	23.37	16.77	43.59	40.72	83.10	78.28
		(4.75)	(10.84)	(7.27)	(9.16)	(23.34)	(22.47)	(4.61)	(14.55)
Environment	Companies	90.31	35.55	90.31	52.32	539.80	86.95	539.80	85.62
		(10.60)	(47.74)	(10.60)	(57.12)	(18.40)	(47.08)	(18.40)	(42.67)
	Percent Cut-off	13.50	40.38	13.50	45.39	80.67	79.57	80.67	82.92
		(1.59)	(36.97)	(1.59)	(35.95)	(2.53)	(20.64)	(2.53)	(12.63)
Human Rights	Companies	606.14	97.12	606.14	96.63	606.14	94.78	606.14	94.78
		(28.67)	(49.35)	(28.67)	(49.04)	(28.67)	(48.25)	(28.67)	(48.25)
	Percent Cut-off	90.59	90.52	90.59	90.66	90.59	91.05	90.59	91.05
		(4.10)	(8.81)	(4.10)	(8.59)	(4.10)	(7.21)	(4.10)	(7.21)
Product	Companies	87.04	17.54	87.04	31.97	524.12	84.82	524.12	83.95
		(13.37)	(11.12)	(13.37)	(34.92)	(31.80)	(45.61)	(31.80)	(41.41)
	Percent Cut-off	13.02	29.54	13.02	33.79	78.35	72.09	78.35	79.41
		(2.03)	(26.09)	(2.03)	(26.08)	(4.84)	(18.25)	(4.84)	(9.69)
Combination 1	Companies	35.75	3.74	81.32	10.32	178.85	26.14	330.51	48.46
		(8.98)	(3.17)	(10.31)	(7.00)	(29.08)	(15.14)	(52.73)	(26.45)
	Percent Cut-off	5.34	4.64	12.15	8.17	26.68	21.99	49.47	45.65
		(1.33)	(3.71)	(1.54)	(3.41)	(4.13)	(5.50)	(8.13)	(9.41)
Combination 2	Companies	30.33	2.65	72.76	9.33	155.13	25.00	291.18	46.15
		(8.67)	(1.72)	(12.91)	(6.98)	(20.51)	(14.48)	(47.34)	(26.21)
	Percent Cut-off	4.53	4.48	10.87	7.68	23.20	20.04	43.51	41.27
		(1.28)	(4.22)	(1.92)	(3.71)	(3.10)	(5.49)	(4.04)	(9.95)

This table summarizes the average annual number of companies and the average annual percent real cut-off included in each portfolio after screening with the standard deviation below in parentheses. The summary data shown here are from the portfolios created using the high-rated ranking screens, which mirror the low-rated ones in terms of the number of companies screened out because both are inclusive of any ties in rank. In the case of the long-short portfolios, the high-ranking portfolio becomes on-inclusive and the low-ranking portfolio remains inclusive to prevent overlap.

The best-in-class columns (2, 4, 6, and 8) are measured by taking the average number of companies chosen per industry class over the course of each portfoilio's 13 year life span.

B. *Table II*

Non-Industry Specific Negative, Positive, and Combination Screens From 1992 to 2004 (Monthly)

Dependent variable	Portfolio Performance In Excess of the Risk Free Rate					
	Alpha (%)	Market	SMB	HML	MOM	R^2
Portfolio screens:	(1)	(2)	(3)	(4)	(5)	(6)
Negative						
high-rated	0.12**	0.97***	-0.18***	0.00	-0.04***	0.98
low-rated	-0.03	0.95***	-0.14***	0.30***	-0.02	0.81
long-short	0.17**	-0.01	-0.06	-0.13*	0.02	0.00
Community						
high-rated	0.25**	0.87***	-0.25***	0.09**	-0.10***	0.88
low-rated	0.08	1.00***	-0.11***	0.04	0.02	0.90
long-short	-1.91***	-0.02	0.17**	0.29***	-0.04	0.09
Diversity						
high-rated	0.03	0.92***	-0.23***	0.06	-0.03	0.88
low-rated	0.06	1.09***	0.04	0.20***	-0.15***	0.82
long-short	0.70***	0.00	0.05	0.07*	-0.02	0.01
Employee Relations						
high-rated	0.23	1.01***	-0.12**	-0.31***	-0.05	0.84
low-rated	-0.10	1.02***	-0.03***	0.04***	-0.14***	0.89
long-short	-0.55**	-0.01	0.00	0.05	0.00	-0.02
Environment						
high-rated	0.24	0.97***	-0.13**	-0.15**	-0.08**	0.79
low-rated	0.05	0.89***	-0.24***	0.26***	-0.04	0.79
long-short	-0.60**	-0.09	0.19***	0.11	-0.04	0.04
Human Rights						
high-rated	0.10**	0.97***	-0.15***	0.02	-0.03***	0.98
low-rated	0.10	0.94***	-0.20***	0.02	-0.03*	0.94
long-short	-2.08***	0.00	0.16**	0.24***	-0.03	0.04
Product						
high-rated	0.03	1.05***	-0.05	-0.17***	-0.06**	0.86
low-rated	0.09	0.87***	-0.32***	0.18***	-0.05***	0.90
long-short	-1.50***	-0.04	0.15**	0.10	-0.04	0.02
Combination 1						
high-rated	0.25*	0.96***	-0.16***	-0.21***	-0.12***	0.89
low-rated	0.04	0.97***	-0.21***	0.31***	-0.04	0.82
long-short	1.18***	-0.05	0.12**	0.08	-0.03	0.03
Combination 2						
high-rated	0.40***	0.95***	-0.15***	-0.20***	-0.14***	0.87
low-rated	0.02	0.90***	-0.22***	0.29***	-0.13***	0.79
long-short	1.43***	-0.02	0.11**	0.08	-0.01	0.04

This table summarizes the monthly results of regressing mutliple value-weighted portfolios from January 1992 to December 2004. These non-industry specific portfolios are regressed on factor loading variables, and the results include coefficients from the independent factor loading variables, the alpha value (y-intercept), and the adjusted-R2 value. All the portfolios are separated into three different investing strategies: high, low, and long-short. The cut-off is 10%.

***Significant at the one percent level.

**Significant at the five percent level.

*Significant at the 10% level.

C. Table III

Best-In-Class Positive and Combination Screens From 1992 to 2004 (Monthly)

Dependent variable	Excess Portfolio Performance Over Risk Free Rate					
	Alpha (%)	Market	SMB	HML	MOM	R^2
Portfolio screens:	(1)	(2)	(3)	(4)	(5)	(6)
Community						
high-rated	-0.16***	0.20***	-0.02	-0.04**	-0.10***	0.80
low-rated	-0.25***	0.29***	0.01	-0.13***	0.02	0.71
long-short	-1.03***	0.01	0.06*	0.15***	-0.02	0.08
Diversity						
high-rated	-0.25***	0.15***	-0.04***	0.04***	-0.01**	0.76
low-rated	-0.34***	0.28***	0.03	0.08***	-0.04***	0.59
long-short	-0.08***	0.00	0.01	0.02*	0.00	0.01
Employee Relations						
high-rated	-0.21***	0.21***	0.00	-0.07***	-0.06***	0.66
low-rated	-0.24***	0.21***	-0.01	0.76***	-0.03***	0.72
long-short	-0.35***	-0.01	0.02	0.02	0.00	0.01
Environment						
high-rated	-0.22***	0.25***	-0.01	0.00	-0.04**	0.63
low-rated	-0.02	0.39***	-0.04	-0.30***	0.07***	0.68
long-short	-1.90***	0.05	0.12*	0.30***	-0.05	0.07
Human Rights						
high-rated	-0.28***	-0.01***	-0.01*	-0.01	-0.01***	0.92
low-rated	-0.13**	0.22***	-0.02	-0.05***	-0.11***	0.77
long-short	-1.21***	0.03	-0.01	0.00	0.01	-0.02
Product						
high-rated	-0.24***	0.23***	0.00	0.02	-0.03***	0.77
low-rated	-0.24***	0.17***	-0.04***	0.04***	0.00	0.85
long-short	-0.46***	-0.01	0.03**	0.02	-0.01	0.02
Combination 1						
high-rated	-0.28***	0.24***	-0.02	0.02	-0.04***	0.73
low-rated	-0.31***	0.19***	-0.03*	0.04**	-0.03***	0.63
long-short	0.24***	-0.01	0.03**	0.03*	0.00	0.03
Combination 2						
high-rated	-0.29***	0.24***	0.00	0.00***	-0.06***	0.73
low-rated	-0.22***	0.23***	0.06***	0.01	0.03**	0.63
long-short	0.46***	-0.02	0.03	0.02	0.00	0.01

This table summarizes the monthly results of regressing mutliple industry- and value-weighted portfolios from January 1992 to December 2004. These best-in-class portfolios are regressed on factor loading variables, and the results include coefficients from the independent factor loading variables, the alpha value (y-intercept), and the adjusted-R^2 value. All the portfolios are separated into three different investing strategies: high, low, and long-short. The cut-off is 10%.

***Significant at the one percent level.

**Significant at the five percent level.

*Significant at the 10% level.

D. Table IV

Non-Industry Specific and Best-In-Class Positive and Combination Screens for Various Cut-offs From 1992 to 2004 (Monthly Alphas in Percent from Long-Short Portfolios)

Dependent variable	Excess Portfolio Performance Over Risk Free Rate			
	5%	10%	25%	50%
Panel A: Non-Industry Specific	(1)	(2)	(3)	(4)
Community	-0.77***	-1.91***	-1.25***	-0.34***
Diversity	0.16*	0.70***	1.67***	1.05***
Employee Relations	-1.57***	-0.55**	-1.19***	-0.23**
Environment	-0.07	-0.60**	-1.69***	-1.86***
Human Rights	-1.85***	-2.08***	-2.40***	-2.40***
Product	-1.37***	-1.50***	-1.26***	-1.57***
Combination 1	0.04	1.18***	0.92***	0.79***
Combination 2	0.92***	1.43***	1.14***	0.77***
	5%	10%	25%	50%
Panel B: Best-In-Class	(1)	(2)	(3)	(4)
Community	-0.46***	-1.03***	-0.48***	-0.31***
Diversity	-0.05*	-0.08***	-0.01	-0.06*
Employee Relations	-0.49***	-0.35***	-0.49***	-0.28***
Environment	-2.38***	-1.90***	-0.56***	-0.56***
Human Rights	-0.62***	-1.21***	-0.71***	-0.66***
Product	-0.45***	-0.46***	-0.66***	-0.51***
Combination 1	-0.02	0.24***	-0.03	-0.15***
Combination 2	0.24***	0.46***	-0.02	-0.10***

This table summarizes the monthly results of regressing mutliple industry- and value-weighted portfolios with various screening cut-offs. The portfolios are regressed on factor loading variables from January 1992 to December 2004. The only result included is the monthly alpha value (y-intercept) from the long-short portfolios.

***Significant at the one percent level.

**Significant at the five percent level.

*Significant at the 10% level.

E. *Table V*

Non-Industry Specific and Best-In-Class Equally-Weighted Portfolios for Negative, Positive, and Combination Screens From 1992 to 2004 (Monthly Alphas in Percent from Long-Short Portfolios)

Dependent variable	Excess Portfolio Performance Over Risk Free Rate	
	Non-Industry Specific	Best-In-Class
Portfolio screens:	(1)	(2)
Negative	0.98***	
Community	-1.59***	-0.89***
Diversity	-0.86***	-0.35***
Employee Relations	-1.37***	-0.54***
Environment	-0.31	-1.43***
Human Rights	-1.52***	-1.03***
Product	-1.16***	-0.46***
Combination 1	0.50***	-0.14***
Combination 2	0.36**	-0.10***

This table summarizes the monthly results of regressing multiple equally-weighted portfolios on factor loading variables from January 1992 to December 2004. The only result included is the monthly alpha value (y-intercept) from the long-short portfolios.

***Significant at the one percent level.
**Significant at the five percent level.
*Significant at the 10% level.

F. *Table VI*

Non-Industry Specific and Best-In-Class Value-Weighted Portfolios for Negative, Positive, and Combination Screens In Two Periods: 1992-1997 and 1998-2004 (Monthly Alphas in Percent from Long-Short Portfolios)

Dependent variable	Excess Portfolio Performance Over Risk Free Rate			
	1992-1997		1998-2004	
	Non-Industry Specific	Best-In-Class	Non-Industry Specific	Best-In-Class
Portfolio screens:	(1)	(2)	(3)	4
Negative	1.58***		1.33***	
Community	-2.86***	-0.98***	-1.13***	-1.06***
Diversity	0.81***	-0.08**	0.63***	-0.06
Employee Relations	-1.68***	-0.57***	0.39	-0.16***
Environment	-3.70***	-0.86***	1.74***	-2.68***
Human Rights	-3.28***	-1.31***	-1.17***	-1.14***
Product	-3.43***	-0.85***	-0.02	-0.16***
Combination 1	-0.06	-0.07*	2.14***	0.49***
Combination 2	0.32	0.06	2.20***	0.77***

This table summarizes the monthly results of regressing multiple industry- and value-weighted portfolios on factor loading variables over two separate periods: January 1992 to December 1997 and January 1998 to December 2004. The only result included is the monthly alpha value (y-intercept) from the long-short portfolios.

***Significant at the one percent level.
**Significant at the five percent level.
*Significant at the 10% level.

IX. References

Bauer, R., Derwall, J., & Otten, R. (2007). The Ethical Mutual Fund Performance Debate: New Evidance from Canada. *Journal of Business Ethics*, 111-124.

Becker, G. (1957). *The Economics of Discrimination.* Chicago: The University of Chicago Press.

Berry, T. C., & Junkus, J. C. (2012, Dec 04). Socially Responsible Investing: An Investor Perspective. *Journal of Business Ethics*, 707-720. Retrieved Apr 12, 2016

Bruin, R. d. (2013). *Delete row if a specific value exists (VBA)*. Retrieved from Ron de Bruin Excel Automation: http://www.rondebruin.nl/win/s4/win001.htm

Buchanan, L. J. (2011). The Success of Long-Short Equity Strategies versus Traditional Equity Strategies & Market Returns. *CMC Senior Theses*.

Carhart, M. M. (1997). On Persistance in Mutual Fund Performance. *The Journal of Finance*, 57-82.

Child, C. (2015). Mainstreaming and its Discontents: Fair Trade, Socially Responsible Investing, and Industry Trajectories. *Journal of Business Ethics*, 601-618.

CRSP. (n.d.). *CRSP Stock Events - Historical Descriptive Information*. Retrieved Feb 29, 2016, from WRDS: https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/stock_q/names.cfm?navId=164

CRSP. (n.d.). *Data Definitions*. Retrieved Mar 29, 2016, from CRSP: http://www.crsp.com/products/documentation/data-definitions-1

Demos, T. (2010, Jun 11). *MSCI to incorporate ESG risks into indices*. Retrieved from Financial Times: http://www.ft.com/cms/s/0/e60be8ee-74e0-11df-aed7-00144feabdc0.html

Dow, C. G. (1998). *"GROWTH" Versus "VALUE" INVESTING*. Retrieved Apr 08, 2016, from Dows: http://www.dows.com/Publications/growth_versus_value_investing.htm

Eurosif. (2010). *European SRI Study 2010.* Paris: Eurosif. Retrieved Apr 10, 2016, from http://www.eurosif.org/publication/european-sri-study-2010/

Fama, E. F., & French, K. R. (1993). Common risk factors int he returns on stocks and bonds. *Journal of Financial Economics*, 3-56.

French, K. (2015, Dec 31). *Detail for 10 Industry Portfolios*. Retrieved from Kenneth R. French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_10_ind_port.html

French, K. (2016). *Changes in CRSP Data*. Retrieved Mar 26, 2016, from Kenneth R. French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/changes_crsp.html

French, K. (2016, Feb 29). *Description of Fama/French Factors*. Retrieved Apr 08, 2016, from Kenneth R. French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

French, K. (2016, February 29). *Detail for Monthly Momentum Factor (Mom)*. Retrieved Apr 08, 2016, from Kenneth R. French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

French, K. (2016, Feb). U.S. Research Returns Data (Downloadable Files).

Geczy, C., Stambaugh, R. F., & Levin, D. (2005, Oct). Investing in Socially Responsible Mutual Funds. *SSRN Electronic Journal*. Retrieved from Social Science Research Network.

Gompers, P. A., Ishii, J. L., & Metrick, A. (2003). Corporate Governance and Equity Prices. *Quarterly Journal of Economics*, 107-156.

Hines, T., & Bordelon, B. (2016, Apr 04). *Finding and using financial data at Princeton University*. Retrieved from Stocks - USA - Data - Daily: http://libguides.princeton.edu/finance/StocksUSA#slgbox161 1231

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 15-36.

Investopedia. (n.d.). *Growth Stock*. Retrieved Apr 12, 2016, from Investopedia: http://www.investopedia.com/terms/g/growthstock.asp

Jacobs, B. I., Levy, K. N., & Starer, D. (1999). Long-Short Portfolio Management: An Integrated Approach. *The Journal of Portfolio Management*, 23-32.

Kempf, A., & Osthoff, P. (2007, Nov 05). The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management, 13*(5), 908-922. Retrieved Feb 11, 2016, from http://onlinelibrary.wiley.com.ezproxy.princeton.edu/doi/10.1111/j.1468-036X.2007.00402.x/epdf

Khan, A. (n.d.). *Java Excel API - A Java API to read, write, and modify Excel spreadsheets*. Retrieved from JExcelApi: http://jexcelapi.sourceforge.net/

Lane, D. M. (n.d.). *Homoscedasticity*. Retrieved from David M Lane: http://davidmlane.com/hyperstat/A121947.html

MSCI, Inc. (2015, Jun). MSCI ESG KLD STATS: 1991-2014 DATA SETS Methodology.

MSCI, Inc. (n.d.). *MSCI ESG KLD STATS - Social Ratings*. Retrieved Feb 26, 2016, from WRDS: https://wrdsweb.wharton.upenn.edu/wrds/ds/kld/criteria.cfm?navId=288

Murray, M. P. (2006). Lecture 14: Heteroskedasticity and Serial Correlation. In M. P. Murray, *Econometrics - A Modern Introduction.* Pearson Addison-Wesley.

Newey, W. K., & West, K. D. (1987). A SIMPLE, POSITIVE SEMI-DEFINITE, HETEROSKEDASTICITY AND AUTOCORRELATION CONSISTENT COVARIANCE MATRIX. *Econometrica*, 703-708.

Princeton Data and Statistical Services. (n.d.). *DS400 Historical Holdings (Inception through November 7, 2008)*. Retrieved Feb 29, 2016, from DSS Princeton: http://dss.princeton.edu/cgibin/catalog/search.cgi?studyno=3092

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

Schueth, S. (2003). Socially Responsible Investing in the United States. *Journal of Business Ethics*, 189-194.

turtle44. (2008, Feb 07). *Excel Training / Excel Dashboards Reports*. Retrieved from Ozgrid: http://www.ozgrid.com/forum/showthread.php?t=84661

von Wallis, M., & Klein, C. (2014, Oct 28). Ethical requirement and financial interest: a literature review on socially responsible investing. (E. Dockner, Ed.) *Business Research*, 61-98.

Wadsworth, J. (n.d.). Lecture 18. Autocorrelation & Dynamic Models. London: Royal Holloway, University of London.

PLEDGE:

This paper represents my own work in accordance with University regulations.

X ______________________________