# Corporate Social Performance and Financial Performance

*Evidence from stakeholder perceptions*

**Abstract**

I investigate the corporate social performance – corporate financial performance (CSP – CFP) link using data on over 800 companies during 2003-2013. In order to define the relationship as cleanly as possible, I differentiate between frontier policies and civil foundation policies. I also add a reputational factor to add some context to CSP. My tests show a clear negative link between CSP and CFP. I also find that stakeholders are largely indifferent between frontier policies and civil foundation policies and that corporate base reputation is ignored when assessing an increase in CSR. Stakeholders assess CSR superficially as they focus on high CSR ratings and are unable to differentiate between true social outperformers and highly rated companies.

# Content

## 1. Introduction

There has been ample research investigating the link between corporate social performance (CSP) and corporate financial performance (CFP). The relation between the two is called the CSP-CFP link or CSP-CFP puzzle as no clear unambiguous effect has been found to date. Some scholars find positive effects (Orlitzky , Schmidt & Rhynes, 2003; Griffin and Mahon, 1997; Cox, Brammer & Millington, 2004; Hillman & Keim, 2001) whereas others find convincing evidence of a negative CSP-CFP relationship (Auperle and Van Pham, 1989; Geczy, Stambaugh & Levin, 2003; Brammer, Brooks & Pavelin 2006). The subject sparked my interest as it combines two fundamentally different concepts. Large multinational businesses that act in the interest of their shareholders, yet undertake policies that benefit everyone but the shareholders. Finding a way to connect the wealth increase of all these other stakeholders to financial gains for shareholders, would mean finding reason for shareholders to encourage social behaviour in businesses. This is why I became interested in the subject, as finding financial benefits of CSP would essentially mean that companies should serve society as a whole and through this optimize serving their owners.

Therefore, I first decided to write my bachelor's thesis on the subject. I investigated the effect that corporate social responsibility (CSR) as measured by KLD ratings had on market risk and downside market risk. I found higher market risk for companies that had low to zero levels of CSR compared to high levels of CSR. The study however didn't use any control variables and focussed solely on risk as a measure of financial performance. Another shortcoming of my bachelor thesis was the blind trust placed in KLD ratings to deliver social outperformers. I was surprised to find in my portfolio of social outperformers an unusual amount of companies in the oil-, fast-food- and agrochemical industries. These are companies that were unlikely to be social outperformers as their core business constituted an unusually high burden to society. Since an investor would not be likely to gain utility from the CSR position of a large oil company, investment decisions regarding these corporations were unlikely to be motivated by their high CSR score. Therefore any effects in financial performance linked to CSR was unlikely to be caused by the variation in CSR levels. The issue with CSR scores as a measure of social performance is in the KLD rating system. The rating system works with a simple tick the box system for certain company policies and does not measure circumstantial factors that could influence stakeholder perception.

Unsatisfied with my discovery I decided to write my master's thesis on the same subject, but I was going to take a completely different approach. In this paper I investigate the corporate social performance (CSP) – corporate financial performance (CFP) relation. Some scholars use CSP and CSR interchangeably, but CSR refers more to the company policies ( as for example measured by the KLD

scores) whereas social performance refers to the results of these actions (Frederick, 1994). The results of company social policies are changes in the perceptions of the various stakeholders that interact with the company. I therefore differentiate between company policies that are likely to influence stakeholder perceptions positively and those that are unlikely to influence stakeholder perceptions positively and may even have a negative influence. This adaption to the KLD database allows me to use the vast KLD database on social responsibility, but relate financial effects to actual social performance instead of only ratings.

Another important addition is that I reintroduce corporate base reputation. Each company in my sample is assigned a good, neutral or bad reputation based on news sentiment in articles regarding CSR or ethics. Adding corporate reputation to CSP places CSP in perspective and allows me to differentiate between for example philanthropy and greenwashing. This enables me to define the relationship between CSP and CFP even more clearly and to approximate stakeholder perception.

Also I use various measures of financial performance to assess the link the redefined CSP measure has to financial performance. Focus on one single measure or combining risk and returns, without regard for the stand-alone variables results in the loss of valuable information. By using both conventional measures of risk as well as measures of downside risk I ensure that I incorporate all relevant measures of financial performance in my research to map possible links between CSP and CFP.

Based on the vast body of academic literature on the subject I expect an unambiguous relation to exist between CSP and CFP. Using my two adaptations to redefine how we measure CSP and using various measures of financial performance, this leads to the following research question:

> Does there exist a relationship between Corporate Social Performance and Financial Performance when both are measured effectively?

The remainder of this thesis is devoted to answering this question and is structured as follows: in section 2 the theoretical framework provides an overview of relevant academic research and section 3 describes the data sources used as well as descriptive statistics on the data. Section 4 specifies the methodology applied to construct the variables and section 5 will show the results of the empirical study. Lastly, section 6 discusses the results, concludes this paper and provides limitations and suggestions for further research.

## 2. Theoretical Framework

In over 30 years of research on the relationship between corporate social performance (CSP) and financial performance (FP) no unambiguous conclusions can be drawn on the relationship between the two. Various meta-studies conducted conclude on a very small but positive relationship between CSP and FP (Griffin & Mahon, 1997; Orlitzky, Schmidt, & Rynes, 2003). However for every study finding a positive relationship, there exist some that find a negative relationship or no (significant) relationship at all. Therefore the debate is still out on the existence and the sign of the relationship.

A positive relationship is upheld by the most commonly supported "instrumental stakeholder theory" First described by Freeman (1994) as the stakeholder approach where the firm functions in a multi-stakeholder environment and its objective should be to manage the interests of all stakeholders. Jones (1995) forms the "instrumental stakeholder view". In the instrumental stakeholder view the level of corporate social responsibility (CSR) enhances the relationship between various important stakeholders and the firm. Enhanced ties with various stakeholders that provide key resources to the company, increases productivity and financial performance. As shown by Brammer, Millington, & Rayton (2007), higher CSR leads to more satisfied employees, attracting/retaining better employees and therefore increases overall productivity. Sen & Bhattacharya (2001) provide evidence of this relationship as their research show that individual-specific factors (such as personal preferences) and company-specific factors (such as which policies of CSR entailed) are key determinants of consumers opinions on CSR . Furthermore, Cox, Brammer, & Millington (2004) provide evidence for the relationship to be existent for investors and Renneboog, Ter Horst & Zhang (2011) show that consistent with the instrumental stakeholder view and multi-attribute utility of investors Bollen (2007), funds flows to Socially Responsible Investment (SRI) funds are less  sensitive to past performance.

The most prominent critique on the stakeholder approach may very well stem from Jensen & Meckling's (1976) "agency theory". Modelling the costs of the "separation of ownership and control", these costs increase with the immeasurability of the agents performance. The multi-stakeholder approach on CSR combined with the agency theory could render a negative relation between CSR and financial performance. As argued by Tirole (2001), nearly any action of the firm's management might be justified in light of the interest of any of the stakeholders. Therefore the multi-stakeholder theory may give rise to agency costs, as management accountability to any specific stakeholder drops and therefore immeasurability of management performance increases. It is the failure to provide a trade-off between the various stakeholders that causes the destruction of firm value and reduction of social welfare (Jensen, 2001).

Another argument for the existence of a negative relationship can be found in Friedman's 1970's article "The social responsibility of business is to increase its profits". Disregarding the discussion whether corporations should engage in philanthropy or the resources and decision belong with the shareholder, actively engaging in CSR is costly. CSP is mainly achieved by either making costs to excel or prevent in some areas or incurring opportunity costs of foregone business opportunities of not excelling or by preventing in others.

The "institutional stakeholder view" is in sharp contrast with the "agency theory" and the (opportunity) costs borne by corporate social performers. To support the theoretical paradox, many scholars found empirical evidence that supports either a positive relationship CSP-FP (Orlitzky, Schmidt & Rhynes, 2003; Griffin and Mahon, 1997; Cox, Brammer & Millington, 2004; Hillman & Keim, 2001) or a negative CSP-FP relationship (Auperle and Van Pham, 1989; Geczy, Stambaugh & Levin, 2003; Brammer, Brooks & Pavelin 2006) It is this contrast that explains the many ambiguous research results and lack of academic consensus on the matter. In addition there is a large body of research that finds no relationship at all or an insignificant relationship at best (Rennenboog, Ter Horst and Zhang 2008; Bauer, Koedijk and Otten 2005, Hamilton, Jo and Statman 1993). The empirical evidence of insignificant relationships was first predicted by Ullman (1985). According to Ullman it would be impossible to statistically verify the relationship because of the many intervening variables.

So one explanation for the observation of no unambiguous CSP-FP link is the fundamental contrast between positive and negative effects of increased CSR. The aim of this study is however not to add to the vast pile of inconclusive evidence, but to examine other possible reasons for the unobserved relationship. One explanation for an unobserved relationship maybe the misspecifications of the variables or consequently the measures of these variables. First, I will discuss the specification of CFP and later the specification of CSR.

The most widely used measure of CFP are market-based measures of a return such as stock returns. For an extensive overview of the various studies that use market based measures versus accounting based measures of financial performance I refer to Margolis and Walsh (2003). The authors provide an overview of which measures of FP are used in 127 studies. Benefit of having a market-based measure of FP is mostly that it is forward looking. As Cox et al. (2004) argue, most benefits of CSR accrue in the long run rather than the short run. However critiqued for containing also broader market information, a forward looking measure is most likely to capture these benefits.

Whereas most studies concentrate on finding a link between the level of stock returns and CSP, some use accounting based measures such as ROA, ROE or EPS (Auperle et al., 1985; Guenster

et al., 2005). As accounting based measures are backward looking and internally oriented measures they are more suited for measuring past performance. They are useful in investigating the bi-directionality of the relationship but less in determining the influence CSP has on FP. According to the good management theory (Waddock and Graves, 1997) companies that have the time and resources available to manage properly their CSR, will in most cases be capable to perform financially. Another theory, claiming similar connections is the "slack resources" theory (Preston & Bannon, 1997). Slack resources theory suggests that when corporations have excess resources they are more likely to engage in CSR. Acknowledging some bi-directionality of the relationship, this study has no further interest in debating the directionality of the CSP-FP link and has no use for backward looking measures. I will therefore not use accounting based measures of CSP.

However, as anyone who read Markowitz's (1952) article on portfolio optimization knows, it is not only the level of returns that is relevant in measuring financial performance. Variance of said returns should not be neglected, though the research in this particular area is very thin. One of the first studies examining the relationship between risk and CSR to some extent was McGuire, Sundgren and Schneeweiss (1988). The authors found an insignificant negative relationship, but used Fortune's ratings of reputation, which (see below) is not an appropriate measure of CSR. The most extensive study on the relationship between risk and CSR must be the recent study of Oikonomou, Brooks and Brammer (2012). The authors use Kinder-Lydenberg-Domini (KLD) social ratings and have a sample of about 7000 firm year observations. The conclusion of the study is that risk is weak yet significantly negatively related to CSR.

The instrumental stakeholder theory predicts that the ties of a company engaging in CSR with its important stakeholders are less contingent upon financial performance as they are with companies not engaging in CSR. Therefore, when financial performance is weaker i.e. in economic downturn, stakeholders of the high CSR company are less inclined to break these ties. This is an adaption of what Oikonomou, Brooks and Brammer (2012) refer to as the wealth-protective effects of CSP and what was the main argument for my bachelor thesis. Both Oikonomou, Brooks and Brammer and myself found a significant negative link between CSP and FP as measured by market risk and downside market risk

Acknowledging the varying results when more in depth measures of financial performance is employed, this study used returns, risk and risk adjusted returns as measures of financial performance so any misspecification of financial performance is avoided.

However it may also very well be possible that there has not been established an unambiguous CSP-FP link due to a misspecification of CSP rather than of FP. CSP is a rather broadly definable concept and finding a construct to measure this proves difficult.

First off, afore defining CSR it is necessary to discriminate between CSR and CSP. Both are central concepts of the literature on the same subject and are sometimes used interchangeably. However, as described by Fernando (1994) CSR relates more to principles and is the collection of corporate actions undertaken in this area whereas CSP is related to the results of this actions. Combining Fernando's insight with the instrumental stakeholder theory we can conclude that CSR refers to actions taken by the firm and CFP to the results achieved by these actions with various stakeholders. In this definition, measures of CSR do not always measure CFP.

In past literature scholars have continuously tried to produce a better measure of CSR. Among the first measures of CSR employed were governmental pollution indices (Bragdon & Marlin, 1972; Sharfman & Fernando, 2008; Salama, Anderson & Toms, 2009). Indices such as from the Council of Economic Priorities rank corporations on their control of pollution. A benefit is that the rankings are likely provided by independent agencies. However, by definition CSR is a multidimensional construct and pollution control ranking, measures a single environmental strength.

Other scholars used content analysis of financial reports (Anderson & Frankle, 1980). Content analysis does allow measuring CSR along all its dimensions, but lacks objectivity. The definition as how to define CSR dimensions may be subjective, and more importantly the data source (financial reports) are likely to be positively biased. Therefore content analysis measures what a company reports, rather than how it performs.

Another commonly used measure of CSR are reputational surveys. Various surveys have been used over time, but most used are the Fortune magazine reputational surveys. (McGuire et al., 1988). A benefit of reputational surveys is that it measures CSP, so outcomes of CSR, rather than CSR itself. However, Fortune reputations are comprised of several financial factors as well and it is the financial factors that dominate the ratings (Fombrun and Shanley, 1990).

And then there is a large quantity of research using CSR as a plug by investigating the difference in financial performance by Social Responsibiltiy Investing (SRI) mutual funds and regular mutual funds (see Rennenboog et al., 2008 for a meta-analysis). SRI funds apply positive screens, only selecting firms that perform above a certain threshold, or negative screens, exempting firms that have certain characteristics. When using the difference between SRI-funds and regular mutual funds as a measure of CSR, this essentially allows the scholars to circumvent quantifying the largely qualitative concept

of CSR. Despite this benefit, the methodology has a high risk to assign any difference in financial performance to the level of CSR/CSP of the stocks in the SRI funds. Even though this difference may arise from for example manager stock picking skills or a reduced investment universe at the portfolio level.

However the vast majority of the more recent studies use KLD ratings (see *inter alia*, Hillman and Keim, 2001, McWilliams and Siegel 2000, Oikonomou, Brooks and Brammer, 2012). Independently measuring CSR over 50 areas in over 3000 companies, this is the most complete measure of CSR available. Huge benefits of the KLD rating systems are the relative objectivity and simplicity. Even though KLD ratings are the best measure of CSR around, there are some flaws. For example Wood (1991) argues that measures of CSR should measure outcomes and should therefore measure CSP. The effectiveness of the measured policies is completely ignored by KLD.

Evaluating the literature review there are several conclusions to be drawn on which to build further research. First of all, there are sound theoretical explanations for both positive and negative CSP-FP links although the meta-analyses favour a positive link. Second, FP should be measured along various measures of FP in order to draw any conclusions. And third, we should develop a measure closer to CSP that is able to measure the outcomes of CSR to some extent. If CSP is measured more effectively I expect that there will be a clear CSP-FP link. This translates into the following research question:

> Research question: Does there exist a relationship between Corporate Social Performance and Financial Performance when both are measured effectively?

There are two major flaws in measuring CSP through the KLD ratings on CSR policies. The first one is that a policy in one of the areas that comprise CSR, is not per se a policy that influences CSP. The essential difference follows from the definition of CSR. A very complete definition of CSR is given by Shahzad and Sharfman (2015). The authors define the concept of CSR based on two essential papers on the definition of CSR (Carroll 1979, Wood 1991) as "a multifaceted construct encompassing the outcome of those voluntary actions taken by corporations that go above and beyond what is legally or economically necessary". A more widely used and older definition of CSR is the one given by McWilliams & Siegel (2001): "actions that appear to further some social good, beyond the interests of the firm and that which is required by law".

Building on the instrumental stakeholder theory: the theoretical relationship exists only because stakeholders gain some sort of utility from engaging in business with the company. Some company policies in areas that comprise CSR such as corporate giving, are however futile in the greater scheme of things and unlikely to yield stakeholders any utility. In order to measure effects hypothesised by

the instrumental stakeholders theory, I therefore must define what goes beyond what economics and laws require. Since what doesn't, shall not impact stakeholder utility positively.

Similar arguments were made by Martin (2002), who sketched a transitional framework, The Virtue Matrix, where policies start out as frontier policies and end up in the civil foundation by either widespread economic acceptance (for example worker treatment by Ford) or are translated into legislation. Frontier policies can be defined as those areas of CSP that are not yet widely accepted and not yet translated into legislation. This corresponds with the definition of CSP to be "above and beyond what is required by economics or law". Civil Foundation policies are defined as what is widely economically accepted and therefore possibly expected or required by law. If we rephrase this, it implies that not every policy in the KLD database and in KLD defined areas contributing to the KLD CSR rating, should be used as a measure of CSP. It is frontier policies, not CF policies that companies get credited for by stakeholders according to Martin. Therefore, using CF policy scores as an indicator of CSP must by definition yield ambiguous results.

The second major flaws of KLD ratings is the use of a binary system (Mattingly, 2015). Every policy either passes a preset cutoff point and scores a 1 or does not make the point and scores 0. The database contains less information than when more continuous measures are employed, but more importantly  the use of this binary rating system does not allow for the necessary relative ranking of the various Strengths required to produce a valid aggregate measure (Ruf, Muralidhar and Paul, 1998).

Discriminating between the relative effects of various CSR policies is especially necessary when the effects on financial performance of the different CSR policies have opposite signs. The negative sign of certain CSR strengths may follow from a practice known as "greenwashing". Even though the term suggests only environmental factors it has been coined to describe the situation where companies appear to be socially responsible but are insincere. Frynas (2005) found that oil companies focus rather on building public facilities such as schools and hospitals, than address the social issues core to the nature of their business. Yoon et al. (2006) provide evidence that the effect on stakeholder perception may be depending on whether or not the company's CSR efforts are sincere and similar reasoning is provided by the study of Godfrey (2005) that argues that efforts to entail in CSR motivated to favourably influence stakeholder perceptions, rather than broader social welfare have negative reputational consequences.

Next to confirming my hypothesis that greenwashing is harmful to company reputation even though they are measured as CSR strengths, the research of Yoon et al. (2006) and Godfrey (2005) provides us with another valuable insight: It is stakeholder perception of the company CSR, that is the

outcome or CSP of the CSR. This leads us to the third major flaw in using KLD ratings to measure CSP: the absence of any stakeholder perception. Following the findings of Frynas, I can state that it is very strange to assume that a policy to build schools has the same effect on stakeholder perception for American Airlines as it has for British Petrol following the 2010 accident with Deepwater Horizon. In this reasoning there are countless examples illustrating the inequality of the effect of one particular KLD strength between various companies.

It was this flaw that triggered my interest in redefining CSR. When using the KLD ratings for my bachelor thesis research on the link between CSR and FP, I isolated a group with the highest KLD ratings. This group seemed to contain more companies that were active in controversial industries than the lower scoring groups. This was according to the KLD ratings supposed to be the forefront of CSR yet contained large oil companies, investment banks and pharmaceuticals that one would not expect to score this high. This shows that the KLD measures should be adapted if they are to contain any real measure of CSP.

## 3. Data

In order to measure CSR, I have used the Kinder, Lydenberg and Domini (KLD) STATS database. KLD is a social rating agency. The ratings assess a company's CSR by investigating the company policies regarding multiple qualitative issues. These issues are split into positives (Strengths) and negatives (Concerns) and measured independent of one another. The individual Strengths and Concerns are grouped into "Areas" along a particular stakeholder which they affect: Community, Governance, Diversity, Employee, Environmental, Product and Human Rights. Because of the broadness of the dataset, the KLD dataset can be applied to assess CSR as the multidimensional construct it is according to the academic literature (Waddock, 2003). All Strengths and Concerns are measured with binary scores for each individual Strength or Concern. Since I will only use Strengths for this paper, I will now solely focus on Strengths.

The KLD database started in 1991 only reviewing the 400 companies of the Domini Social 400 and has experienced quite some changes ever since. Some minor changes in Strengths' definitions before 1995, but more important are the dataset expansions. In 2001, the KLD dataset was expanded to include the 1000 largest US companies and in 2003 it was expanded again, now to include the US largest 3000 US companies. As of recently, KLD became part of Morgan Stanley Capital International (MSCI) and along with a new owner came some methodological changes. The various Strengths were revised in both 2012 and 2013. In this revision, some Strengths were discontinued and grouped into a new composite Strength, others were simply discontinued or merged. Also entire new strengths were added to dataset.

However potentially harmful for the continuity of our measurements, the changes in methodology are also a benefit of the KLD database. As CSR is a multidimensional construct concerning company policies effecting an ever changing society as a whole, definitions change and therefore methodology should be dynamic as well. Also, apart from the recent changes, for the larger part of my sample period KLD has entailed consistent application of objective criteria.

The dataset is assembled from multiple sources, *inter alia* annual surveys from company representatives, financial statement analysis and expert assessment of annual reports. The multi source data gathering contributes to the objectivity of the KLD dataset.

The KLD dataset is also the most widely used dataset and has become the measure of CSR for some of the most important academic papers in the field of CSR (Among others: Waddock and Graves, 1997; Hillman & Keim, 2001; McWilliams and Siegel ,2001). Bakker, Groenenwegen and den

Hond (2005) acknowledge the importance of converging standards for the progression of the literature in this field. It is therefore of even more importance to question this standard.

All market data and company fundamentals have been acquired from the CRSP and COMPUSTAT databases. Both databases have an extensive amount of data that more than suits my needs for this study. The widespread use of CRSP and COMPUSTAT data in academic literature affirms the quality of the data.

For the methodology I have often sought external statistic confirmation of intuition and theories. All data for this purpose was acquired from independent sources. Preferably governmental or semi-governmental organizations to ensure its independency. Among others these are the US Bureau of Labour Statistics (BLS), Occupational Safety and Health Association (OSHA) and Committee Engaging Corporate Philantropy (CECP).

[Factiva]

**3.1 Descriptives and Mutations**

Descriptive statistics on return characteristics of the dataset are given in the Table 1 (Equally Weighted) and 2 (Value Weighted). In both tables descriptive statistics on the whole dataset from 2003 to 2013 are mentioned under the column "ALL". All the other columns contain information on the various subsets or "portfolios" of the "ALL" set. The first two columns differentiate along the KLD ratings between high CSR scores "High Rating" and low CSR scores "Low Rating". These portfolios contain the companies that scored high or low based on KLD's unadapted ratings. Columns 3 and 4 contain descripitive statistics on two portfolios differentiated along the Frontier versus Civil Foundation Methodology.

Columns 5 to 10 contain descriptives on highly specialised portfolios made by adding a reputational factor and are used to verify various theories concerning the CSR-CFP link with my adapted dataset.

| Equally Weighted Monthly Returns | Mean | Std Error | Median | Std Dev. | Var | Kurtosis | Skewness | Min | Max |
|---|---|---|---|---|---|---|---|---|---|
| ALL | 0.015 | 0.005 | 0.023 | 0.056 | 0.003 | 3.791 | -0.398 | -0.224 | 0.238 |
| High CSR | 0.011 | 0.005 | 0.013 | 0.053 | 0.003 | 3.768 | -0.421 | -0.212 | 0.221 |
| Low CSR | 0.010 | 0.005 | 0.014 | 0.056 | 0.003 | 5.205 | -0.269 | -0.207 | 0.264 |
| High FP | 0.013 | 0.006 | 0.019 | 0.073 | 0.005 | 3.948 | -0.328 | -0.269 | 0.318 |
| High CF | 0.012 | 0.004 | 0.023 | 0.050 | 0.003 | 2.920 | -0.946 | -0.211 | 0.144 |
| High FP - Good Rep. | 0.011 | 0.005 | 0.017 | 0.056 | 0.003 | 5.127 | -1.057 | -0.263 | 0.204 |
| High CF - Good Rep. | 0.010 | 0.005 | 0.011 | 0.053 | 0.003 | 5.136 | -0.194 | -0.219 | 0.249 |
| Low CSR - Bad Rep. | 0.011 | 0.004 | 0.016 | 0.050 | 0.002 | 3.879 | -0.403 | -0.198 | 0.205 |
| Mid CF - Bad Rep. | 0.010 | 0.004 | 0.016 | 0.049 | 0.002 | 3.348 | -0.342 | -0.189 | 0.203 |
| Mid CF - Neutral Rep. | 0.016 | 0.005 | 0.025 | 0.059 | 0.003 | 2.956 | -0.374 | -0.220 | 0.241 |
| Good Rep. | 0.012 | 0.005 | 0.015 | 0.054 | 0.003 | 3.449 | -0.260 | -0.203 | 0.232 |

**Table 1: Return Characteristics of subsets based on Equally Weighted Monthly Returns**

Discriminating between equally weighted and value weighted monthly returns does not seem to produce all to different descriptive statistics. Only in the highly specialised portfolios a small difference exists, but this was to be expected as highly specializing means compromising on numbers.

| Value Weighted Monthly Returns | Mean | Std Error | Median | Std Dev. | Var | Kurtosis | Skewness | Min | Max |
|---|---|---|---|---|---|---|---|---|---|
| ALL | 0.014 | 0.004 | 0.017 | 0.041 | 0.002 | 1.855 | -0.562 | -0.152 | 0.123 |
| High CSR | 0.010 | 0.003 | 0.016 | 0.038 | 0.001 | 1.330 | -0.570 | -0.126 | 0.099 |
| Low CSR | 0.016 | 0.004 | 0.017 | 0.051 | 0.003 | 2.184 | -0.351 | -0.186 | 0.187 |
| High FP | 0.009 | 0.004 | 0.014 | 0.043 | 0.002 | 2.826 | -0.521 | -0.165 | 0.151 |
| High CF | 0.011 | 0.003 | 0.015 | 0.039 | 0.001 | 1.407 | -0.563 | -0.132 | 0.108 |
| High FP - Good Rep. | 0.012 | 0.004 | 0.010 | 0.047 | 0.002 | 1.787 | -0.108 | -0.160 | 0.177 |
| High CF - Good Rep. | 0.012 | 0.004 | 0.012 | 0.042 | 0.002 | 1.560 | -0.434 | -0.139 | 0.119 |
| Low CSR - Bad Rep. | 0.015 | 0.005 | 0.012 | 0.060 | 0.004 | 1.674 | -0.251 | -0.184 | 0.195 |
| Mid CF - Bad Rep. | 0.016 | 0.004 | 0.018 | 0.046 | 0.002 | 1.580 | -0.641 | -0.172 | 0.118 |
| Mid CF - Neutral Rep. | 0.016 | 0.004 | 0.022 | 0.050 | 0.002 | 4.278 | -1.021 | -0.221 | 0.164 |
| Good Rep. | 0.012 | 0.004 | 0.013 | 0.043 | 0.002 | 1.616 | -0.292 | -0.149 | 0.145 |

**Table 1: Return Characteristics of subsets based on Value Weighted Monthly Returns**

Furthermore, Table 3 contains an overview of the amount of companies (n) that have been included in each portfolio.

| Number of companies | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| High CSR | 48 | 33 | 58 | 64 | 70 | 70 | 67 | 133 | 134 | 88 | 92 |
| Low CSR | 135 | 142 | 155 | 146 | 142 | 138 | 138 | 103 | 109 | 92 | 98 |
| High FP | 38 | 39 | 42 | 40 | 41 | 43 | 44 | 58 | 59 | 38 | 37 |
| High CF | 70 | 34 | 59 | 81 | 97 | 99 | 96 | 197 | 200 | 140 | 156 |
| High FP - Good Rep. | 36 | 34 | 35 | 37 | 28 | 28 | 27 | 34 | 34 | 29 | 30 |
| High CF - Good Rep. | 28 | 30 | 37 | 31 | 38 | 36 | 37 | 64 | 51 | 59 | 61 |
| Low CSR - Bad Rep. | 38 | 39 | 31 | 32 | 29 | 31 | 44 | 27 | 28 | 33 | 29 |
| Mid CF - Bad Rep. | 36 | 31 | 37 | 37 | 45 | 45 | 39 | 58 | 59 | 54 | 53 |
| Mid CF - Neutral Rep. | 41 | 42 | 49 | 32 | 43 | 48 | 35 | 82 | 58 | 35 | 42 |
| Good Rep. | 75 | 77 | 78 | 74 | 76 | 78 | 122 | 121 | 118 | 120 | 121 |

**Table 2: Number of companies (n) per subset**

In total the dataset contains unbalanced panel data including a KLD score, a reputation factor, yearly accounting information and monthly market information on 871 different companies over 11 years. This yields a total of over 90.000 firm month observations, which is largely sufficient for my research purposes.

## 4. Methodology

### 4.1 Measuring CSR levels: KLD scores

As explained in my theoretical framework a reason for not observing the hypothesised link between CSP and CFP may be misspecification of the CSR measure. If the aggregate measure of CSR aggregates contrary effects, these effects may counteract one another and leave us with a net effect that is unobservable. This paper hypothesises that the cause of the contrary effects may well be found in how stakeholders perceive certain policies. Ultimately it is their perception of the information provided that influences their actions, not the information (CSR) provided by companies itself. The way KLD measures CSR, subdivides CSR into different areas and subsequently aggregates along these areas does measure CSR but does not measure how stakeholders may perceive the CSR efforts. Therefore, adaption of the database is required to allow aggregation along possible stakeholder perceptions.

To define how policies will be viewed, I need to distinct between policies that are likely to influence stakeholder perception positively and those that may not influence stakeholder perception at all or even have a negative influence (Godfrey, 2005). In order to do so, I build on a transitional model developed by Martin (2002) known as the "Virtue Matrix". Martin describes a model that can be used to assess whether CSR policies are on the "frontier" or in the "civil foundation". The civil foundation (CF) is "an accumulation of customs, norms, laws and regulations". A policy is therefore either civil foundation by choice (customs and norms) or by compliance (laws and regulations). Frontier policies (FP) on the other hand, are those that are not yet in the civil foundation as they are not required to be so by norms and customs or by regulations. As Martin describes frontier policies, they are somewhat revolutionary and pioneering. After a while they can become civil foundation policies as acceptance grows or legislation is made.

The complete definition of CSR given by Shahzad and Sharfman (2015): "a multifaceted construct encompassing the outcome of those voluntary actions taken by corporations that go above and beyond what is legally or economically necessary" can as well be used to identify the difference between frontier policies and the civil foundation. The important part of the definition is here "that go above and beyond what is legally or economically necessary". As that what is "legally or economically necessary" corresponds to what qualifies as civil foundation by either compliance and by choice. Therefore, all what goes above and beyond that what is legally or economically necessary may be classified as frontier policy and can be expected to influence stakeholder perception positively. Policies that are legally or economically necessary are part of the civil foundation and are less likely to influence stakeholder perception positively as stakeholders are unlikely to see this as

CSP and if stakeholders view some civil foundation policies as hypocritical the influence on stakeholder perception may even be negative.

In table 5 I have reported all KLD Strengths included in my dataset and classified those as either CF, FP or decided to exclude them. As this is very soft qualitative data finding external validation for my arguments is not always possible. Acknowledging this, most variables have some source of external validation or distinction can be made by sound reasoning. Table 4 provides an overview of the totals and table 5 the classification of the individual variables.

| KLD Area | CF | FP | Excluded |
|---|---|---|---|
| Community | 6 | 1 | 0 |
| Corporate Governance | 1 | 4 | 2 |
| Diversity | 4 | 4 | 0 |
| Employment | 5 | 3 | 2 |
| Environment | 1 | 6 | 0 |
| Human Rights | 0 | 2 | 0 |
| Product | 0 | 1 | 3 |
| Total | 17 | 21 | 7 |

Table 4: Overview of Strengths classifications to CF and FP per area and total

Table 5: KLD strength classification into civil foundation (CF) and frontier policies (FP)

| KLD Strength | Class | Motivation |
|---|---|---|
| Community | | |
| Charitable giving:<br><br>Company has donated 1.5% of NEBIT for a period of 3 years to charitable cause | CF | Probably the best known policy of corporate pro-social behaviour, the amount required for KLD to consider this policy a strength is fairly low. According to the CECP reports in 2013, 25% of US companies donated over 1.95% of NEBIT. Also, the median donation was around 1% over the whole sample period. In sample-percentages for 2010 to 2011 average 12%. This is lower than expected, but a steep rise from 2008 2009, which indicates cyclicality in charitable giving levels. The low in-sample percentage can also be the result of the KLD criterion requiring a consecutive period of three years. Even though the KLD threshold resulting in 12% seems to be tighter than expected, the CECP reports indicate that there are many companies that employ a higher level of charitable giving. The added criterion of three years in a row at the KLD levels of giving I expect not to change shareholder perception to revolutionary. I therefore classify this as a civil foundation (CF) policy. |
| Innovative giving:<br><br>Innovative support for non-profit organisations and stimulation of workplace giving by employees. | CF | Innovative giving has two aspects: innovative support for non-profit organisations and workplace giving stimulation. Innovative support for non-profit organisations is hard to define, but extrapolating the findings on charitable giving, support for non-profits is likely to be widely implemented.<br>On the employee giving stimulation there is very convincing evidence for it to be CF. CECP reports in 2013 that 86% of companies researched had some sort of employee matching gifts program of which 66% was |

| | | all year round. These percentage fluctuate around 90% for earlier years. The generalness of this policy and mass application in US corporates make it required by economics and a CF policy. This is supported by the in-sample percentages of around 35%. |
|---|---|---|
| Charitable Giving International: <br><br> Of all donations at least 20% is outside of the US | CF | The international character is unlikely to change the nature of the charitable giving policy. The CECP reports confirm this as in 2013 of all donations, 22% was international. In earlier years this figure fluctuated around 15%, making 20% far from a best in class cut-off point. Also CECP shows that this figure was much higher for companies with revenues of over 30% from outside the US. This shows the direct link with economics and the motivation for giving to some extent. The in-sample percentage for this is in 2008-2009 around 6%, this is at the same level as charitable giving in these crisis years. Similar to charitable giving, in-sample percentages are less informative because of cyclicality and the lack of observations for recent non-crisis years. The main consideration for classifying Charitable Giving International is the same as with Charitable Giving, apparently (based on the CECP reports) it is very widely applied just under the KLD threshold. Making the KLD threshold is unlikely to change stakeholder perception to a frontier policy. I therefore classify this policy as civil foundation. |
| Support for Housing: <br><br> Company stimulates housing for economically disadvantaged. E.g. is a member of National Equity Fund or the Enterprise foundation | FP | Housing for the economically disadvantaged is a rather specific cause for companies to give to. That is: the philanthropic nature of giving to a cause unrelated to the underlying business is higher with such a specific cause. The CECP reports show that over the sample period the average given to Community and Economic Development is only 14% of all donations. Of this 14%, only a small part is aimed at housing support. The highly specialist nature of this policy makes it likely a frontier policy as support for housing is hardly the norm in society. Donations to stimulate housing for economically disadvantaged is also not required by law, confirming the FP classification. This is supported by in-sample statistics of around 3% for years 2008-2009. |
| Support for Education: <br><br> Support for primary or secondary education | CF | Donations to support education also have a specialist nature and can in no way be linked to the underlying business of most corporations. These are philanthropic characteristics that indicate FP. However, the CECP reports show that over the sample period around 30% of all donations were destined for education. This is a very high percentage, making this policy more of a commodity than a revolutionary FP or even at least required above and beyond economics. This is supported by the, for crisis years, high in-sample percentages of around 10% over 2008-2009. |
| Volunteer Programs: <br><br> Company has a notably strong volunteer program | FP | Closely linked to the matching gifts programs and stimulation of workplace giving by employees. CECP reports that 86% of the researched companies had a volunteer program where at least 6 hours of volunteering on company time could be done. This makes Volunteer Programs a very common policy and likely to be part of the CF. However, the in-sample percentages are only 4% indicating that for a volunteer program to be "notably strong" it should well be above 6 – |

| | | |
|---|---|---|
| | | hours. As the KLD threshold is very strict it is likely to measure only those examples that can be seen as frontier policies. I therefore classify Volunteer Programs as FP. |
| Community Engagement:<br><br>Company has a notable community engagement program in areas where the firm has its major operations. | EXCL. | A combined variable that has been created after the transition to MSCI to replace various community strengths, among others charitable giving and volunteering. Since the thresholds are undefined for this policy and the individual components are not uniformly classified as either FP or CF, I exclude this variable. |
| **Governance** | | |
| Limited compensation:<br><br>CEO compensation < $500,000 and outside directors <$30,000 | FP | Average CEO compensation for publicly listed companies runs in to the millions making limited CEO compensation a very rare event within this sample. Having executive compensation at such low levels is truly a revolutionary frontier policy. In sample percentages of around 2% support the rarity of this policy. |
| Ownership Strength:<br><br>Owns 20% or is owned for 20% by another KLD noted company that scores well on social strengths. | EXCL. | Variable is excluded since it is impossible to determine a classification for the strengths on which the parent or subsidiary scores well. Also impossible to externally verify importance as the variable exists only within a KLD dataset. Furthermore, in sample statistics of close to 0 show that classification of this variable is unlikely to influence results. |
| Transparency Strength:<br><br>Company is very effective in reporting on social and environmental reporting measures | CF | This is a widely applied policy as this policy is both a transparency strength and a way of communicating social performance to stakeholders. A study by KDPaine shows that 80% of companies researched extensively report social and environmental progress on their website. Most do this as a dedicated sub-domain and some as an independent website. Although it is not required by US law yet, it is now required in the European Union by law. This is a clear sign that in Western culture this is CF. This is supported by average in sample percentages around 20% for 2010-2013. |
| Political Accountability:<br><br>Can a company represent company and shareholder interests in politics, but with broad approval and transparency policies? Therefore, avoid any legal and reputational risks accompanied with political support | FP | The *Citizens United* Supreme court case was to provide more transparency regarding corporate influence in politics. Legislation building on this was the Disclose act, a bill for more transparency and accountability regarding corporate donation to political campaigns, but was defeated in the senate. Recent scandals with the presidential campaign funding of Hillary Clinton and failure or unwillingness to turn this into legislation indicate that a sound, accountable policy for political spending is not yet part of our civil foundation. Certainly it is also not required by economics or the norm in society. I indicate this as a frontier policy. My findings are supported by the low in-sample percentages of around 2%. |
| Corruption & Political Instability: | FP | Corruption and political accountability are very closely linked. In Accountability it is support of politics to obtain political influence. |

| | | |
|---|---|---|
| Evaluates whether company policies effectively avoid operational disruptions, loss of market access, property destruction, sabotage, demands for bribes due to corruption. | | Corruption is the one step further, directly achieving or preventing an action of officials by payment, intimidation or other leverage. Transparency International ranks USA as the 16th least perceived corrupt country in the world, but does state that approximately almost 72% of US citizens experienced a rise in corruption over the period 2007-2010.It is scandals like LIBOR manipulation that influence the perception of corruption in corporate America. As the United states are relatively not a very corrupt country, having an outspoken anti corruption policy in the US is unlikely to be seen as a frontier policy. However, nearly all companies in this sample are multinationals and having an anti corruption policy is a frontier policy in a lot of countries. As this is not required by law or economics, I deem this a frontier policy. This is supported by the in-sample percentage of 8% |
| Financial System Instability:<br><br>How well does a company manage its systemic risk in financial markets? | FP | Only included as of 2012 this variable is a response on the 2008 financial crisis. This score can only be obtained by companies exposed to systemic risk i.e. financial institutions. There have been set out all kinds of regulatory frameworks to reduce systemic risk in financial institutions (Solvency and BASEL) that going above and beyond what is required by these frameworks in a competitive market is a definite frontier policy. This is supported by in-sample statistics of 10%. |
| Public Policy Support:<br><br>Company supports public policies that has beneficial effects for: environment, employees, consumers or communities. | EXCL. | Excluded from the sample. This is a very soft criterion on which no external validation could be found. The in sample percentage show negligible levels that will not affect results when in- or excluded. |
| Diversity | | |
| CEO Diversity:<br><br>CEO is female or member of a ethnic minority | FP | Diversity is a concept that has been around for a while and in itself has largely become part of the civil foundation. At the executive level this is rather a different story. Of all S&P 500 companies, a meagre 23 have a female CEO and only 6 CEOs are Afro-American. This in a country where over half the residents is female and 13.2% is Afro-American this is hardly a reflection of society. Therefore at executive level, it is still revolutionary to have diversity which makes this Frontier Policy. This is supported by in-sample percentages of around 4%. |
| Promotion:<br><br>Company has policies for fair promotion chances | CF | Having fair promotion chances is required by law (several civil rights acts) which would make it automatically CF. In sample percentages that are consistently nearly 30% confirm the widespread adoption of equal promotion policies. |
| Board Diversity:<br><br>Board contains 33.3% women and ethnic | FP | Similar arguments apply here as to CEO Diversity. Executive level diversity is still very low. Zweigenhaft and Domhoff (2011) have extensively studied diversity on corporate boards of Fortune 500 companies. They show severe underrepresentation for white females, |

| | | |
|---|---|---|
| minorities | | both black males and females, both Asian-American males and females and both Latino males and females. With 74.4% of all corporate directors being white males. Only 25.6% is to be divided among white women, and all other ethnic groups. This is not laid down by law, and very seldom met at 33.3% of the board being else than white males. In sample percentages are close to 8%  and thus supporting the FP classification. |
| Work/Life Benefits:

Policies on for example maternity/paternity leaves, sabbaticals, weekly hours etc. | FP | In the U.S. work/life balances are not highly regulated by law. Some companies like Google and Apple are famous for taking care of their employees and have policies that redefine our perception on how employment should be. On the other hand are standards very low. For example legally there is no paternity leave and legal maternity leave is only 12 weeks. In addition job security after maternity leave is only 56%. In sample percentages are around 15%, not excluding either of the classifications, so for this criterion I build solely on theory. As standards are still low and outperformers influence the way society defines employment, this policy can be viewed as on the frontier of change and therefore I believe an FP classification is justified. |
| Women and Minority Contracting:

At least 5% of subcontracting at minority or women owned business | FP | As subcontracting policies are less visible to external stakeholders, they have a more philanthropic character by definition. Female entrepreneurship has been on the rise and grows at 1.5 times the national average growth in new businesses. This policy goes above and beyond what is required by law and or economics and can therefore be classified as a frontier policy. In-sample percentage of around 10% support the FP classification. |
| Employment of Disabled:

Company any notable policy that increases employment of disabled persons | FP | Legislation on hiring disabled persons has been around since the early nineties. However, The Bureau of Labour Statistics (BLS) reports that in 2012 and 2013 unemployment of disabled was still three times higher than for workers without disability. This implies that employment of the disabled is still a pressing issue. In-sample percentages of around 5% indicate that the KLD definition is quite strict. Combining the in sample percentages with the BLS figures supports an FP classification. |
| Gay & Lesbian Policies:

Company has particular policies that improve Gay and Lesbian equality. Most importantly is employee partner health benefits. | CF | Gay rights acceptance has made significant progress in recent years. In 2015 the supreme court ruling in the Obergefell-Hodges case implies same sex marriage is legal in U.S., therefore putting an end to spousal benefits discrimination and confirming the CF status of this policy. Also in earlier years this was already in the CF. In 2013 the supreme court already ruled that any tax reduction that was valid for spousal health benefits in opposite sex marriage, would be valid for same sex marriages. However not all states allowed same sex marriage, therefore constraining the effect of this ruling. The Human Rights Campaign or HRC, researched the issue in 2004 and 2006 and found in these years that of all nearly 10,000 companies offering domestic partner health benefits for both years, 95% of these policies applied to same sex spouses as well. Making this, also in earlier sample years, a definite CF policy. This is supported by in-sample averages of over 45% in recent years. |

| | | |
|---|---|---|
| Employment of Underrepresented groups:

Includes the now discontinued employment of the disabled. But is not limited to that. Evaluates also diversity concerning race and age of the workforce. | CF | Employment of the disabled I classified as frontier policy. But what about employment of the various other underrepresented groups? The BLS reports for the year 2012, 2013 and 2014 that labor force participation rates are highly similar for the various ethnic minorities as compared to US whites. However, these are the stats for bullish years. Unemployment rates for blacks and Hispanics was much higher in 2008-2010 during the financial crisis. The report shows that when white unemployment rises Black and Hispanic unemployment mimics this but is amplified. According to the report this may be due to a tendency for these races to work in jobs that know higher unemployment and require less education, but may also be due to discrimination. Given the BLS statistics I classify this CF policy in good years, but FP in bad. Since it is only measured like this in the last four years of the sample, the variable receives a CF classification. In-sample percentages of over 30% in the last four years support the CF classification. |
| **Employee Relations** | | |
| Union Relations:

Company has maintained a policy to treat the unionized workforce fairly | CF | Unions have been around for a very long time and are subject to various macro-economic transitions. For example, the economy becoming more service-oriented and the rise of low-cost countries in Asia have reduced the impact of unions. The BLS shows that union membership was as high as over 1/3 of the workforce in 1945, but steadily decreased to a mere 11 percent in 2014 (1998 = 13.9% and 1980s = 24.1%). The long history of interacting with unions has made this a CF policy. In sample percentages are around 6%, but this is in line with expectations based on the BLS statistics and is no reason for not classifying this as CF. |
| Cash profit sharing program:

The company has one and has recently made distributions to a majority of its workforce | FP | Since the distribution must be made to a majority of the workforce, executive stock options are not included in this area. Also there is a strict distinction between a bonus (individual performance based) and cash-profit sharing (company performance based). The National Compensation Survey (NCS) from the BLS shows that on average for the sample period only 5% of respondents received a cash-profit sharing. Cash profit-sharing among employees is not a widespread business model, but empowers all layers of society. It is safe to say that cash profit sharing is above and beyond what is required by economics or law and therefore classified as FP. In-sample averages are around 10% supporting the FP classification. |
| Employee involvement:

Company strongly encourages involvement and ownership through: stock options or stock ownership for a majority of the workforce. Also involvement in | FP | According to the National Center of Employee Ownership (NCEO), development of Employee Stock Ownership Plans or ESOPs first saw light in 1974. Ownership can be given through actual stocks or stock options. According to the National Compenstation Survey (NCS) of 2012, 8% of private sector employees had company stock options. However, this does not mention whether the majority of a workforce owns these stock options, or only management. NCEO shows a steady increase of ESOP participants over years 2002-2013, but also shows that only 3% of |

| | | |
|---|---|---|
| management decisions. | | ESOPs are with public companies. This shows that the definition of KLD is fairly strict compared to the status quo. In-sample percentages are around 15%, not strictly indicating either of the classifications. The criterion that ownership is for the majority of the workforce I believe deserves an FP classification. |
| Employee Pension Benefit Strength:<br><br>Has a very notable pension strength either a well funded defined benefit program, or a notably well individual benefit program 401(k) | CF | Pension plans are heavily regulated under and required by U.S. law. The pensions system is very old and completely part of the Civil Foundation. However, pension plans have changed over time. U.S. Treasury researched trends in the era from 1977-2007 and found that most notable was the change from defined benefit plans to defined contribution plans (401(k)). Where defined benefit plans, which are most beneficial employees, declined over the period by 31%, 401(k) increased nearly 400%. In defined benefit plans the weight of payments and the risk of pension asset values are on employer side. That weight is shifting to more voluntary contribution structures of the 401(k). However, changes in recent years are foremost results of changing demographics and low interest rate climate. Having a proper pension plan has been around for a long time and can be considered a CF policy. This is supported by in-sample percentages of nearly 20% in crisis years 2008/2009. |
| Health and Safety Strengths:<br><br>Company has very notable health and safety plans | CF | The Occupational Safety and Health Act of 1970 lead to the establishment of Occupational Safety & Health Administration (OSHA), allowing government officials to control working conditions at private companies, across all 50 states.<br> OSHA describes the trend in workplace safety as from the 1970s workplace fatalities have declined by 66% and injuries by 67%.The workforce has meanwhile doubled, implying a significant improvement in safety & health programs has resulted in reduced injuries and fatalities. The modern working environment differs essentially from the work environment 50 years ago. However, as in developed countries human life has become increasingly more valuable, in many countries it has not. This specific strength concerns only employees and not supply chain though and as this encompasses mostly US and developed country workers, this is a definite CF policy. In-sample percentages of around 20% support this classification. |
| Supply Chain Policies, Programs & Initiatives<br><br>Measures the strength of company policies that are designed to monitor and manage human and labor rights performance of its suppliers and contractors. | FP | In the mid-nineties the world took notice of sweat shop labour that was used by US based apparel manufacturer NIKE. In these sweatshops often under aged workers worked under appalling conditions. Emmelhainz & Adams (1999) show that pressure from public opinion, forced apparel companies to adjust, improving overall supply chain conditions in the apparel industry. However, there still exist many stories of extortion and bad labour conditions in low-wage countries. Although exact statistics are hard to find for these developing countries, International Labour Organization (ILO) reports alarming statistics on average work weeks increasing to over 46 hours. Supply chain policies have drawn attention since the nineties but it is mostly what the public eye catches that is improved as supply chains remain largely |

| | | intransparent (Maloni & Brown, 2006). Having proper supply chain policies goes well above and beyond what is required by law and economics. It is therefore classified as a frontier policy. In-sample averages of 4% support the FP classification. |
|---|---|---|
| Compensation & Benefits<br><br>Companies that have noteworthy employee compensation and benefit programs. | EXCL. | A very broad category that is to include discontinued Pension Plan Benefits (CF) and some previously unspecified policies. The mixed nature of this policy makes this policy impossible to externally verify. In addition it only has values for years 2012-2013. The low amount of observations and unverifiable nature of this variable have led to exclusion of the variable. |
| Employee Relations<br><br>Measures companies that provide collective bargaining, involvement programs or actively measure employee satisfaction. | EXCL. | A very broad category that is to include both collective bargaining (already covered in union relations), involvement programs and actively measuring employee satisfaction. In addition the variable only has values for years 2012-2013. The in-sample statistics show 10% for 2012 and 0% for 2013. Also levels are expected to be much higher for measuring employee satisfaction. These inconsistencies, the nature of the variable and the lack of observations have led to the exclusion of this variable. |
| Professional Development:<br><br>Captures companies that provide excellent training and development programs. | CF | This variable is rather a strategic business choice with positive externalities than a philanthropic policy. That is, for most companies this is required by economics as in many businesses well trained employees are the path to creating and maintaining competitive advantages. Association for Talent Development makes and annual report called the *state of the industry.* Although not publicly available it does reveal spending of an average 1200 dollar on training per employee in 2014. This is an 20% increase from crisis lows in 2008. Also a year on year increase in hours of training is noted.<br>The high levels of spending on employee development and the necessity of the policy for many companies makes this a CF policy. In-sample averages over 2012 and 2013 were as high as 60% supporting the CF classification |
| Human Capital Management:<br><br>Measures company's ability to attract, retain and develop human capital as well as avoid labor unrest due to poor job satisfaction based on their provision of benefits, training and development programs and employee engagement | CF | Similar arguments apply to this variable as to training and development. Attracting and retaining satisfied human capital is a strategic business choice and has the positive network externality of educating the workforce. This is required by economics for many companies and therefore CF. In-sample percentages of upto 30% for 2012 and 2013 support the CF classification. |

| Environment | | |
|---|---|---|
| Beneficial Products and Services:<br><br>A substantial part of company revenues come from products or services believed to have a relatively positive influence on the environment. | FP | Producing products with an environmentally friendly nature is by definition not required by economics or by law. For example, producing toxic chemicals is not illegal, the production process may be regulated and require permissions, but the nature of the product is not (exception for nuclear energy, guns and drugs).<br>This example emphasizes the philanthropic nature of the variable. Where many policies can be used to window dress, making green products or producing green services isn't one of them. Therefore I deem Beneficial Products and Services FP. In sample averages of 5% for earlier years and 10% for later years support the FP classification. |
| Pollution prevention:<br><br>Company has notably strong pollution prevention programs with BOTH emissions reductions and toxic-use reductions. | FP | Khanna and Damon (1999) plea that there is widespread recognition for firms implementing pollution prevention programs. The authors conclude that over the period 1993-1999 pollution significantly decreased as a result of a pollution prevention program of the US Environmental Protection Agency in the chemicals industry. Vidovic and Khanna (2005) review on this work and state that the program did not actually beat the independent trend that was partaking in the chemicals industry. Their research in the matter shows two things: one, there is an independent trend starting in the late 1980's causing pollution prevention and two, US government tried the first pollution reduction program in early 1990's. So having a pollution prevention program is to some extent required by law and economics as this issue has received very much press over time. The distinction here is a notably strong pollution prevention program. As in-sample percentages of only 4% show, this definition is very strict indicating that companies that score on this strength go well above and beyond what is required by law or economics when it comes to pollution prevention. Therefore, this policy is classified as an FP policy. |
| Recycling:<br><br>Company uses a lot of recycled materials in the production process, or is a major recycling company. | FP | Corporations like Coca-Cola and WalMart have set ambitious targets for the use of recycled materials (25% by 2015), but are among the very few companies that have done so (The Economist, 2015). This shows that including lots of recycled materials in the production process is a frontier policy. To recycling as a business model similar arguments apply as to Beneficial Products and Services. This implies that Recycling is a definite FP. In-sample percentages of a mere 2% support the FP classification. |
| Clean Energy:<br><br>Companies make formal statements to support reduction of greenhouse gasses or efforts to reduce climate change | CF | CERES report on sustainability show that in 2013 43% of Fortune 500 companies set energy targets. This goes to show just how widespread this policy is. It is however not required by law and doesn't directly influence any particular stakeholder. It is also not directly required by economics and climate change remains a pressing issue. However the criterion of making a formal statement, does not seem to be very strict. In-sample percentages of over 40% show that even though in theory this may go above and beyond what is required by economics, in practice it is an easy pr-win. This implies a CF classification. |

| | | |
|---|---|---|
| Management Systems:<br><br>Company has a commitment to environmentalism through a voluntary certification program such as ISO 14001 | CF | ISO 14001 Environmental Management System Certification is carried out by the International Organisation for Standardization or ISO. According to ISO in the USA in 1999, 636 firms were certified. This steadily increased to 6000 in the year 2002, around which it fluctuated for the years 2002-2012.<br>The sheer amount of companies that have ISO 14001 certification makes this a CF policy. In-sample percentages of above 30% support the CF classification. |
| Water Stress:<br><br>Company has active policy to manage and contain water stress | FP | CERES' 2012 report on water stress shows that drought, floods and contamination have significant impact on both human welfare and corporate financial performance. In a study of 100 fortune 500 companies in industries related to heavy water use such as mining and beverage industries they found that disclosure on water management systems is growing, but still limited and that there still is a lack of quantitative data and performance targets.<br><br>Another CERES report in cooperation with Sustainalytics emphasizes that figures on water stress are even worse when including supply chains in the methodology. The very low statistics show there is no wide acceptation of water stress policies as of yet and corporations fail to set quantifiable targets. Setting water stress targets goes well beyond requirements by law or economics and is therefore FP. The in-sample percentages of 7% are a clear indicator of the FP classification. |
| Biodiversity & Land Use:<br><br>Programs that manage the risk of damaging fragile ecosystems and protect biodiversity. | FP | Typically and traditionally an area where governments and super nationals are active. It is therefore mostly governments, that develop Biodiversity Action Plans (BAP). Sample statistics are really low since only few companies are confronted with biodiversity issues. Establishing programs that protect biodiversity however completely goes beyond what is required by law, economics or pleasing a certain group of stakeholders. Supported by in-sample percentages of an almost negligible 1%, I classify this as FP. |
| Raw Material Sourcing:<br><br>Company has a policy that promotes transparency and reduces environmental impact of raw materials used | FP | Hard to find external validation on. Raw materials sourcing is very similar to supply chain management, but focuses more on the environmental aspect as opposed to social factors. Also there is similarity with the beneficial products and services<br>As similar reasoning applies Raw Material Sourcing can be classified an FP policy. In-sample percentages of around 3% support this classification. |
| Human Rights | | |
| Indigenous people strengths<br><br>Company has relationships with indigenous people near current or proposed operations that respect | FP | According to the UN, relationships between industrial corporations an indigenous people have always been bad. The worst practices are however found in extractive industries. Particularly gas, oil, mining and foresting companies have bad relationships with indigenous people. This is because the lands inhabited by indigenous tribes often contain large deposits of natural resources, but indigenous people having significant cultural binding with those lands. Making extraction of the natural resources difficult without conflict. I can't assess legislation on |

| | | |
|---|---|---|
| land culture sovereignty etc. | | the subject because of the international character of the subject, but in 2007 General Assembly of the UN accepted the UN resolution that is now known as the Declaration on the Rights of Indigenous Peoples. The widespread malpractice and the fact that having proper relationships with indigenous people is not the norm in society, make a strong policy on this a FP. In-sample percentages of around 5% agree with the above. |
| Labour Rights Strengths: Company has outstanding transparency on overseas sourcing disclosure and monitoring, has particularly good union relationships outside U.S. or has initiated labour-rights initiatives that are deemed innovative | FP | A way of cutting costs is outsourcing low-skilled work to low cost countries (LCC). These countries do not excel at the work, but lack any legal framework protecting workers from extortion. Some countries like India and Pakistan, do not have a minimum age requirement for work effectively tolerating child labour. In Bangladesh, the average workweek is 46 hours. These are just examples of how little protection is offered to workers by the legal system. Having any initiatives to improve life for overseas employees in manufacturing, is in no way required by law in most LCCs. And Bhattacharya shows that differences in labour costs between high and low cost countries range from 2 to 20. That is, in some countries labour is a staggering 20 times cheaper than in the US (Bhattacharya et al. 2004). Any policies that companies voluntarily engage in to improve labour rights, working conditions or unionizing are not required by law or economics and therefore considered frontier policies. In-sample percentages of mostly under 1% fully support the explanation above. |
| Product | | |
| Product Quality: Company has a long term, well developed quality programme. | EXCL. | This variable is excluded. Producing high quality goods and services is a differentiation strategy that has positive externalities (fewer waste, high customer satisfaction etc.). However, I will not test whether diversification as a strategy is linked to financial performance since the answer is known. Therefore this variable is excluded. |
| R&D + Innovation The company is a leader in its industry for research and development (R&D), particularly by bringing notably innovative products to market. | EXCL. | This variable is excluded. R&D + Innovation is a strategic business decision with positive externalities (progression). I will not try to relate R&D intensity to financial performance. |
| Benefits to Economically Disadvantaged. The company has as part of its basic mission the provision of products or services for the economically disadvantaged. | FP | Same arguments apply here as those provided under Beneficial Products and Services and Recycling. The nature of the product is pro-social and this goes well beyond what is required by legislation or economics. In-sample percentages of around 2% support the FP classification. |

| Access to Capital | EXCL. | Access to capital is a variable that could ensure continuity, which is desirable from a social perspective (job security, welfare increase etc.) but it is just like product quality linked directly with financial performance. Since I don't want to test whether corporations with access to multiple sources of capital are funded more expensive or cheaper this variable is excluded. |
| --- | --- | --- |

**4.2 Measuring prior reputation: Factiva Reputation Indication**

A crucial factor in assessing the impact of corporate pro-social actions as measured by the KLD scores is the reputation of the actor. Carroll (2010) argues that in order to properly assess the effect of any corporate action on company reputation it should be regarded in the light of the existing reputation. Furthermore, Lyon and Cameron(2004) have shown that the perception of bad corporate media coverage was evaluated in the light of existing reputation. Bae and Cameron (2006) research the way stakeholders assess corporate giving. Their findings confirm my intuition that stakeholders assess corporate philanthropy in the light of the existing corporate reputation and grow suspicious of company motives for CSR when a company has a bad reputation.

These research results can all be supported by the instrumental stakeholder approach. Following the stakeholder approach to CSR, it is ultimately the perception that the stakeholders have of the company that determines whether stakeholders want to engage in more favourable mutual relationships with a company based on CSR. A reputation is defined by Fombrun, Gardberg & Barnett (2000) as "a cognitive representation of a company's actions and results that crystallizes the firm's ability to deliver valued outcomes to its stakeholders ". Paramount in this definition are the cognitive representation, or the perception of a company's actions, and the acknowledgement of a multi-stakeholder environment.

The abstractness of perceptions and the broadness of the multi-stakeholder environment, cause corporate reputation to be a variable that is hard to quantify. Some important corporate actions and results that influence the company stakeholders' cognitive representation of the company are: financial performance, CSP and executive leadership. Because of this broadness of the reputational concept and since I am only interested in the perception of CSR, I cannot use traditional measures of reputation. So a required characteristic of the measure of reputation employed is in my case that it measures the CSP part of the reputation and doesn't measure the financial performance or executive leadership factors. I have therefore constructed a reputational indicator by measuring news sentiment in newspaper articles using the Factiva database and the Factiva expert search function. Using news sentiment as an indication for corporate CSR reputation is not a perfect measure, but it has some important characteristics. First of all, combining news media with my CSP measure encompasses nearly all information available to stakeholders in forming perceptions regarding CSP. News articles sentiment measures not actual, but possible stakeholder perceptions of the company as news media are an important resource of information on a company. Another very important source of information on company CSP, is company published information, which I captured in the CSP measure, so the combined measure should encompass most of the publicly available information

regarding CSP about a company upon which stakeholders can stake perceptions. Second, it allows me to filter solely on CSR related topics, so a company CSR reputation isn't dominated by financial news. Third, news articles are likely more objective, whereas company published information is possibly positively biased in favour of the company. Furthermore, news articles and other media coverage have been used before in academic literature as a proxy for corporate reputation (Bae and Caroll, 2006).

In order to filter out only the CSR topic related news I have built an extensive search query that encompasses all aspects related to CSR. I have used the extensive KLD descriptions of both Strengths and Concerns to identify topics that may influence stakeholders perceptions regarding the CSR reputation of a company. I have refrained from including all factors, since some tend to dominate in certain industries. The factors I left out were: "fraud" and "lawsuit". Fraud for example should be included as engaging in fraudulent practices negatively influences reputation, but when included it yielded false positives for every company that was a victim of fraud. Financial institutions and especially credit card companies receive a skewed reputation when fraud is included. I also excluded "lawsuits", lawsuits against companies especially by members of the public or the government are a negative influence on corporate reputation. However, in many industries that are heavy in intangibles such as biotech, entertainment or tech, corporate lawsuits such as patent infringements are very common practice and not at all hurting corporate CSR reputation. Including lawsuits would overweight negative sentiment in these industries and was therefore not included. Please note that these terms were not excluded from the search, simply not included. This means that lawsuits do pop-up if they are affecting any of the included search areas. So a lawsuit concerning environmental issues is still included in the search. Further, I have used the Factiva Expert Search tool to filter out positive and negative news sentiment.

As the process required manual adjustment per search it was very time-consuming. Also the Factiva database contains mostly information post 2003. Together with the notion that corporate reputations are sticky (Ang and Wight, 2009) this has led me to use only two observation points over the research period. These observation moments are: year-end 2009 and year-end 2013. As reputation is proxied by the *cumulative* news coverage over a period, the total cumulative coverage up to a certain point in time can be argued to have constructed the reputation at that point. The points 2008 and 2013 are chosen to improve statistical validity of the measure. Ideally, I would have also used the start of the sample period. However, as not all news was electronic, the Factiva database comprised so little observations prior to 2003, that unambiguously assigning a reputation was not possible.

The Factiva Reputation Indicator is set up in such a way that each company was assigned one of three reputation indicators: good, neutral or bad. Table 6 contains the cut-off criteria for assigning any of the reputational indicators. For simplicity, I have assumed that a negative observation is at par with a positive one.

The first cut-off I make is whether the absolute amount of either positives or negatives is sufficiently high to unambiguously assign a reputation. The cut-off point for this is both >=10 articles with positive sentiment (positives) and >=10 articles with negative sentiment (negatives). If there are insufficient article, so either positives or negatives <10, these companies are deemed not eligible and are assigned neutral. Stakeholders are unlikely to have a strong opinion on the CSR of a company that only marginally made the newspapers in 13 years. After the first cut-off, I am left with a set of companies eligible to be assigned a reputation indicator (eligibles). The second cut-off points are based on the relation between the absolute difference and the relative difference. The larger the absolute difference, the lower the required absolute difference to form an opinion. All numbers are the amount of negative news stories divided by the total amount of news stories. So 0.75, implies three quarters of total results are negative and a quarter are positive. If the relative difference is such that there are three times as many negatives (>0.75) as there are positives, the reputation indicator is "bad" regardless of the absolute difference (provided that the company makes the second cut-off). Same goes for positives. However, when relative differences are smaller (0.6 or 0.4), there is an additional requirement of an absolute difference over 5. If the differences are too small the company gets assigned a "neutral" reputation indicator, as positives and negatives cancel out.

| Negatives | Positives | Relative Difference | Absolute Difference | Reputation |
|---|---|---|---|---|
| First Cut-Off | | | | |
| < 10 | < 10 | - | - | Neutral |
| >= 10 | >= 10 | - | - | Eligible |
| Second Cut-Off: Eligibles | | | | |
| | | > 0.75 | - | Bad |
| | | < 0.25 | - | Good |
| | | > 0.6 | >= 5 | Bad |
| | | < 0.4 | >= 5 | Good |
| | | 0.4 - 0.6 | >= 5 | Neutral |
| | | - | < 5 | Neutral |

**Table 6: Determining Factiva Reputation Indicator**

**4.3 Measuring Financial Performance**

Using the adapted dataset I employ various different measures of financial performance. Using more in depth financial measures will increase the possible links between CSP and CFP. For a detailed discussion on which measures to employ, I refer to the theoretical framework. As stated before I will only use market-based measures of financial performance.

All market data are monthly stock returns. The use of monthly returns reduces the amount of observations as compared to daily returns and weekly returns, but over the whole sample yields 132 observations per portfolio nevertheless. According to the Central Limit Theorem, monthly data yields sufficient observations to obtain statistical validity when testing. Daily stock prices are quite noisy and therefore less fit for this research. Weekly data would have actually be the preferred choice, but were not readily available and would have been cumbersome to construct, for very limited statistical gains. The first measure employed is a measure of return and is simply the mean of the returns observed.

The second measure employed is a measure of market risk. This will be the $\beta^{Market}$ figure from the Fama and French three factor model. This is an adaption of the Mosin, Lintner and Sharpe CAPM that is the most widely for modelling financial performance. Using a widely used model enhances comparability to different studies. Set in a Mean-Variance or M-V framework, the model assigns the variation in stock prices to the sensitivity $\beta^{Market}$ of a certain stock to the market returns $E(r_M)$ over and above the risk-free rate $r_f$. This is also referred to as the market risk premium. Since the $\beta^{Market}$ is the sensitivity of a company's stock returns to the market returns it measures a company's non-diversifiable or systemic risk. For any stock or portfolio $i$ of stocks the CAPM can be specified as:

$$E(r_i) = r_f + \beta^{Market}(E(r_M))$$

Equation 1

And $\beta^{Market}$ is calculated as the covariance between the specific stock returns and market stock returns divided by the variance of the market returns.

$$\beta^{Market} = \frac{Cov(r_i, r_M)}{Var(r_M)}$$

Equation 2

The use of the $\beta^{Market}$ figure requires inputs for $r_f$ and for $r_M$. As the market portfolio is unknown and there is no known truly risk free asset I will use commonly employed proxies. For the market returns this is the S&P 500. Although my sample is mainly made up of S&P constituents, I work with

portfolios of companies that comprise only a small portion of the total. Therefore the contrast will be sufficiently large, even though some similarity exists between the two. For the risk-free rate I will use annual interest rates on 3-month U.S. Treasury Bills. Furthermore, the model employs two control variables about which explanation will follow in section 4.4

The third measure employed is the Bawa and Lindenberg β or β$^{BL}$. In 1975 Bawa assessed the portfolio optimization problem in a Mean-Lower Partial Moment or M-LPM. The second order LPM or LPM$_2$ only measures downside volatility as opposed to traditional volatility that measures both up- and downside volatility. The distinction is of vital importance when considering the wealth protective effects of CSR since the value of wealth protection is greatest in downside scenarios. LPM$_2$ is specified in equation 3 where $y$ is the return on a stock and $r_t$ is the threshold rate.

$$LPM_2(r_f, X) = \int_a^{r_f} (r_t - y)^2 dF_x(y)$$

**Equation 3**

In 1977 Bawa together with Lindenberg developed an asset pricing mode similar to the CAPM at large, but set in a M-LPM framework. This model is given by equation 4.

$$E(r_i) = r_f + \beta^{BL}(E(r_M) - r_f)$$

**Equation 4**

In this model β$^{BL}$ is an indicator of the non-diversifiable systemic downside risk. The formula for β$^{BL}$ is given in equation 5. In the Bawa Lindenberg model the threshold is the risk-free rate $r_f$. The construction of the β$^{BL}$ figure is quite similar to the composition of the β$^{Market}$. It is the ratio of the second order Co-Lower Partial Moment CLPM$_2$ (M-LPM equivalent of co-variance) to the LPM$_2$ of the market portfolio.

$$\beta^{LPM} = \frac{CLPM(r_f, r_i, r_M)}{LPM_2(r_f, r_M)}$$

**Equation 5**

$$LPM_2(r_f, r_M) = \frac{1}{T} \sum^T [\min(r_M - r_f, 0)]^2$$

**Equation 7**

$$CLPM_2(r_f; r_M; r_i) = \frac{1}{T} \sum^T (r_i - r_f) \min[(r_M - r_f), 0]$$

**Equation 6**

The last measure of financial performance employed is measuring risk-adjusted returns. This measure combines measures 1 and 2 and can be a simple dummy variable denoted *b* on the relevant portfolio characteristic.

**4.4 Control Variables**

In the following subsection I describe the control variables employed to isolate the effects of CSP. The list is by no means exhaustive but should provide explanation for the majority of return variation, effectively isolating the effects of CSP and reducing the possibility that any observed effects assigned to CSP are merely the result of omitted variable bias.

The two factors added to the CAPM are the two additional factors included in the Fama-French three factor model. The High Minus Low (HML) and the Small Minus Big (SMB).

The HML factor measures the historical difference in returns between value stocks (high book-to-market-value) and growth stocks (low book-to-market-value). The book-to-market ratio is may be related to the risk of financial distress (Fama and French, 1992). Moreover, stocks with a high book to market value are expected to have worse prospects as compared to growth stocks. However, the uncertainty of the prospects of growth stocks is much higher than that of value stocks. Therefore, the riskiness of a stock as well as the level of returns is influenced by the book-to-market ratio.

The SMB factor measures the historical difference in returns between small market capitalization stocks and large (big) market capitalization stocks. This factor includes a size into the model. Size is arguably of influence when assessing financial performance. Large firms tend to be less risky as they are better able to withstand economic downturn than small market value stocks. Small cap firms however, have higher growth potential combined with higher probability of default and variance of returns.

## 4.5 Model Specification

Building on the above I will use the following models to evaluate the CSP-FP link. Model 1 is the Fama-French three factor model used for assessing conventional market risk as measured by $\beta^{Market}$. I run model 1 for each subset to obtain a measure of conventional market risk.

$$E(r_i) = r_f + \beta^{Market}(E(r_M) - r_f) + \beta^{HML}(HML) + \beta^{SMB}(SMB)$$

**Model 1**

The second model I use is an adaption to the Fama-French three factor model to which I add dummy variables for the various  portfolios. To test the various testing hypotheses, I apply this model to the specific subset of the data and regress on one of the dummies of the involved portfolios.  For example to assess the difference in performance between hypothetical portfolios A and B, I run model  1 + dummy for A or B on a subset containing only observations that occur in A or B. This model is labelled model 2 and provides the difference in risk adjusted returns for the test.

$$E(r_i) = r_f + \beta^{Market}(E(r_M) - r_f) + \beta^{HML}(HML) + \beta^{SMB}(SMB) + \beta^{R\&D}(R\&DIntensity) \\ + b^1(Portfolio\ 1)$$

**Model 2**

As testing for differences in means requires a simple student t-test and $\beta^{BL}$ figures are calculated and not estimated, no further regression models are required.

**4.6 Forming Testing Hypotheses**

In my theoretical framework I developed the following research question: Does there exist a relationship between Corporate Social Performance and Financial Performance when both are measured effectively?

I have put in extensive effort to build a database that effectively measures both social performance and financial performance accurately. I will use this specially amended database to answer my research question. Any effects are I expect to be most visible in the extremes. However, also less extreme situations are tested, to increase the study's comparability and to build and extend on existing literature. Also, that way I can test if for this specific period in time my database and methodology allow for better identification of the CSP and CFP link. This is preferred to comparing it simply to existing studies as these cover different timeframes, which may lead to inferring wrong conclusions regarding the effectiveness of my methodology.

The first two research hypotheses build on prior research. Given the large amount of academic evidence provided by the meta-studies (Waddock and Graves, 1997; Orlitzky, Scmidt and Rhynes, 2003) on the CSP CFP link, I expect a small positive or insignificant link between CSP and CFP as measured simply by the aggregate KLD statistic. That is, companies that simply have a high CSR rating (High CSR). H1 tests whether High CSR companies outperform companies that do not have High CSR (Other,. H2 tests whether High CSR outperforms low but non-zero CSR (Low CSR). This translates to the following testing hypotheses:

$H1a$: $\mu_{High\ CSR} > \mu_{OTHER}$

$H1b$: $\beta^{CAPM}_{High\ CSR} - \beta^{CAPM}_{OTHER} < 0$

$H1c$: $\beta^{BL}_{High-CSR} - \beta^{BL}_{OTHER} < 0$

$H1d$: $b_{HI-CSR} > b_{OTHER}$

$H2a$: $\mu_{High-CSR} > \mu_{Low\ CSR}$

$H2b$: $\beta^{CAPM}_{High\ CSR} - \beta^{CAPM}_{Low\ CSR} < 0$

$H2c$: $\beta^{BL}_{High\ CSR} - \beta^{BL}_{Low\ CSR} < 0$

$H2d$: $b_{High\ CSR} > b_{Low\ CSR}$

The third hypothesis builds on the notion that some CSR is more beneficial to stakeholder perceptions than other types of CSR. Using the methodology of Martin (2002), the distinction between frontier policies and civil foundation policies may yield different financial performance when stakeholders generally favour frontier policies over civil foundation policies. Adding to this is the definitional distinction between CSR and CSP that assigns policies that are likely to have more philanthropic motives to the FP. I hypothesise that stakeholders take notice of this and reward companies for FP more than for CF. Therefore I expect companies that have high scores in frontier policies (High FP) to financially outperform companies that score high in civil foundation policies (High CF). This yields the following testing hypothesis:

$H3a$: $\mu_{High\ FP} > \mu_{High\ CF}$

$H3b$: $\beta^{CAPM}_{High\ FP} - \beta^{CAPM}_{High\ CF} < 0$

$H3c$: $\beta^{BL}_{High\ FP} - \beta^{BL}_{High\ CF} < 0$

$H3d$: $b_{High\ FP} > b_{High\ CF}$

The fourth to seventh hypotheses test whether company motives and sincerity ultimately influence stakeholder perception  of CSR. My expectation based on the literature and intuition is that stakeholders punish greenwashing and hypocrisy. This would mean that companies that engage in greenwashing i.e. perform some low to mid level CF whilst having a bad reputation  (Mid CF – Bad) are expected to underperform the rest of the sample (Other). When stakeholders punish hypocrisy this would imply that these greenwashing companies  (Mid CF – Bad Rep.) also underperform companies that don't engage in CSR and have a bad reputation (Low CSR – Bad Rep.) Furthermore I also expect this to hold for the positive counterparts. This means that stakeholders reward sincerity and social outperformance. Therefore I expect companies that score high in FP and have a good reputation to outperform both the rest of the sample (Other) as tested by H6 and companies that have only a good reputation (Good Rep.), yet are no social outperformers. These are the resulting testing hypotheses:

$H4a$: $\mu_{Mid\ CF\ -\ Bad\ Rep.} < \mu_{Other}$

$H5a$: $\mu_{Mid\ CF-Bad\ Rep.} < \mu_{Low\ CSR-Bad\ Rep.}$

$H4b$: $\beta^{CAPM}_{Mid\ CF-Bad\ Rep.} - \beta^{Market}_{Other} > 0$

$H5b$: $\beta^{CAPM}_{Mid\ CF-Bad\ Rep.} - \beta^{CAPM}_{Low\ CSR-Bad\ Rep.} > 0$

$H4c$: $\beta^{BL}_{Mid\ CF-Bad\ Rep.} - \beta^{BL}_{Other} > 0$

$H5c$: $\beta^{BL}_{Mid\ CF-Bad\ Rep.} - \beta^{BL}_{Low\ CSR-Bad\ Rep.} > 0$

$H4d$: $b_{Mid\ CF-Bad\ Rep.} < b_{Other}$

$H5d$: $b_{Mid\ CF-Bad\ Rep.} < b_{Low\ CSR-Bad\ Rep.}$

$H6a$: $\mu_{High\ FP-Good\ Rep.} > \mu_{Other}$

$H7a$: $\mu_{High\ FP-Good\ Rep.} > \mu_{Good\ Rep.}$

$H6b$: $\beta^{CAPM}_{High\ FP-Good\ Rep.} - \beta^{CAPM}_{Other} < 0$

$H7b$: $\beta^{CAPM}_{High\ FP-Good\ Rep.} - \beta^{CAPM}_{Good\ Rep.} < 0$

$H6c$: $\beta^{BL}_{High\ FP-Good\ Rep.} - \beta^{BL}_{Other} < 0$

$H7c$: $\beta^{BL}_{High\ FP-Good\ Rep.} - \beta^{BL}_{Good\ Rep.} < 0$

$H6d$: $b_{High\ FP-Good\ Rep.} > b_{OTHER}$

$H7d$: $b_{High\ FP-Good\ Rep.} > b_{Good\ Rep.}$

The final hypotheses will test whether there exists a stuck in the middle situation when it comes to CSR. Based on Porter's famous typology a company strategy should be either low-cost (Low CSR – Bad Rep.) or differentiated (High FP – Good Rep.), but being stuck in the middle (Mid CF – Neutral) will underperform these two strategies (Porter, 1980). Also the work of Oikonomou, Brooks and Pavelin (2014) shows that stakeholders reward CSR uniformity over mixed forms of positives and negatives. Making the most of my specialised dataset, I will apply their theory to my dataset. This leads to the following testing hypotheses:

$H8a$: $\mu_{High\ FP-Good\ Rep.} > \mu_{Mid\ CF-Neutral}$

$H9a$: $\mu_{Mid\ CF-Neutral} < \mu_{Low\ CSR-Bad\ Rep.}$

$H8b$: $\beta^{CAPM}_{High\ FP-Good\ Rep.} - \beta^{CAPM}_{Mid\ CF-Neutral} < 0$

$H9b$: $\beta^{CAPM}_{Mid\ CF-Neutral} - \beta^{CAPM}_{Low\ CSR-Bad\ Rep.} > 0$

$H8c$: $\beta^{BL}_{High\ FP-Good} - \beta^{BL}_{Mid\ CF-Neutral} < 0$

$H9c$: $\beta^{BL}_{Mid\ CF-Neutral} - \beta^{BL}_{Low\ CSR-Bad\ Rep.} > 0$

$H8d$: $b_{High\ FP-Good} > b_{Mid\ CF-Neutral}$

$H9d$: $b_{Mid\ CF-Neutral} < b_{Low\ CSR-Bad\ Rep.}$

**4.7 Method**

In order to test various aspects of the hypothesised CSP-CFP link I use various subsets (portfolios) of the specialised dataset and compare financial performance between these portfolios. Characteristics along which the subdivision is done are the CSR measure as reported by KLD, CSP measure according to my methodology and reputational factors in line with my methodology.

Using a portfolio structure is of importance as it allows me to differ along ordinal levels of CSP as opposed to absolute levels of CSP. This means that any measurement errors in the absolute values do not automatically take effect in the CSP measure. The measurement of CSP is a quantification of multi-disciplinary, very soft, qualitative data. With this kind of data some measurement errors are inevitable. Therefore, measurement errors will occur either in my methodology or in the KLD methodology. Ranking and choosing ordinal segregation reduces the risk of these measurement errors affecting study results. Observations that lay far away from the cut-off point are largely immune to measurement errors influencing results. The amount of observations that may be prone to measurement errors has now effectively been reduced to those that are situated directly above or beneath the cut-off points. For an overview of the cut-off points used please see table 7. To obtain statistical validity, cut off points are structured such that sample size is over 30 observations per specific portfolio.

| Portfolio | CF | FP | Total | Reputation |
|---|---|---|---|---|
| High CSR | - | - | >=7 | - |
| Low CSR | - | - | 1 | - |
| High FP | <=2 | >=4 | - | - |
| High CF | >=4 | <=2 | - | - |
| High FP – Good Rep. | <3 | >=3 | - | Good |
| High CF – Good Rep. | >=3 | <=2 | - | Good |
| Low CSR – Bad Rep. | <=1 | <=1 | - | Bad |
| Mid CF – Bad Rep. | >=2 | <=2 | - | Bad |
| Mid CF – Neutral Rep. | >=3 | <=2 | - | Neutral |
| Good Rep. | - | - | - | Good |

Table 7

To obtain a measure of CSP I use disaggregate measures with the favourable characteristics of an aggregate measure. That is, I use aggregation along the lines of whether some policies may be perceived positively by stakeholders. Therefore my FP and CF measures are aggregated, but should

be less mixed in terms of stakeholder perceptions. The use of aggregate measures to measure CSP has been widely accepted in academic literature (Waddock & Graves, 1997; Oikonomou, Brooks & Brammer, 2012) and the use of one aggregate measure allows for the uniform evaluation of companies over the complete width of policies that comprises CSR. Also, using individual CSR points show no significant effects for any individual point and the strongest link exists with the aggregate measure (Oikonomou, Brooks & Brammer, 2012). However, the use of an aggregate measure results in loss of information and even more harmful contains effects of opposite signs (Brammer, Brooks & Pavelin, 2006). Therefore, some level of disaggregation is required to obtain an unambiguous measure of CSP. Acknowledging the necessity of disaggregate measure I discriminate between FP and CF, yet aggregate the policies in these categories to obtain my FP and CF aggregate measures. For continuity referred to as FP and CF measures.

As to the statistical tests employed to test the hypothesised relationships and answer the research question, I have used student t-test to test for differences in means. This allows me to test the a. hypotheses. To test for differences in risk adjusted returns (d. Hypotheses), I have used F-tests to test for the significance of the included dummy variables. To test whether the b. hypotheses test statistics are statistically different from zero I have also used F-tests, however when comparing two datasets I could not statistically verify the difference between two test statistics. Test statistics for the c. hypotheses are calculated instead of estimated, therefore unfortunately no statistical tests can be employed to statistically verify the $\beta^{BL}$ figures.

## 5. Results

In this section all results of the aforementioned testing hypotheses are shown and analyzed. Section 6 contains a general overview of the implications of these results as well as a comparison to existing academic literature on the subject. Furthermore, I would like to provide the reader with a quick note on the interpretation of the various tables. In the upper left corner of the tables the labels of the specific testing hypotheses are noted, these correspond with the testing hypotheses formed in the Methodology section. All *a* sub-hypotheses denoted by *H#a* are student t-tests and presented statistics are the result of a student t-test. All *d* sub-hypotheses denoted by *H#d* are multiple regressions on the three factors of the Fama and French three factor model and a Dummy variable for any of the two specific portfolios in that testing hypothesis. The *a* and *d* hypotheses results tables are therefore fairly straightforward to interpret. The *b* sub-hypotheses tables are less straightforward to interpret as it is not possible to test for differences in $\beta^{Market}$ figures estimated from two different populations. As a result, the table consists of two separate regressions, performed on two separate populations. Testing statistics mentioned, therefore do not hold any information on the difference between these $\beta^{Market}$ figures, but rather on the likelihood of them being equal to zero. All regressions used for the b sub-hypotheses used the control variables. All control variables were significant at the 0.01 significance level, but serve no purpose other than preventing omitted variables bias. As the control variables are no indicators of financial performance, they are fairly irrelevant and I therefore do not report statistics on these for the *b* sub-hypotheses . As for the *c* sub-hypotheses denoted *H#c*, the Bawa-Lindenbergh Beta $\beta^{BL}$ is calculated rather than estimated via regression. This means there are  no further statistics to be reported apart from the $\beta^{BL}$ coefficient.

### 5.1 Hypothesis 1 (H1)

The first testing hypothesis tests the relationship between companies that have high levels of CSR (High CSR) as reported by the KLD database and financial performance. This is a very general measure, but provides valuable information on whether investors value CSR rankings such as KLD. H1a shows that there is no statistical difference in average returns between companies with high levels of  CSR compared to the rest of the sample (Other). However statistically insignificant, a difference in annual returns of on average 5% over 10 years is economically significant. The results indicate that companies with high levels of CSR underperform the rest of the sample.

| H1a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High CSR | 0.0102 | 0.0492 | 132 | 257 | -0.7898 | 0.2152 |
| Other | 0.0153 | 0.0568 | 132 | | | |

The results of testing H1b show that the market risk that is associated with companies that have high levels of CSR seems to be lower than the market risk of the rest of the sample. This illustrates the wealth protective effects of CSR. Further confirmation of the wealth-protective effects of CSR comes from H1c that shows that the difference in market risk is still existent when considering downside market risk.

| H1b | β Market | Robust Std. Error | t | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High CSR | 1.0224 | 0.027 | 37.67 | 0 | 0.97 | 1.08 |
| Other | 1.1544 | 0.012 | 96.94 | 0 | 1.13 | 1.18 |

| H1c | $\beta^{BL}$ |
|---|---|
| High CSR | 1.057 |
| Other | 1.177 |

When testing for the difference in risk-adjusted performance it becomes clear that companies with high CSR levels significantly underperform the rest of the sample. The coefficient on the dummy is also fairly high at minus 0.07, implying high CSR companies to achieve annual returns of 7% lower than the other companies in the sample.

| H1d | Coef. | Robust Std. Error | t-value | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.1372 | 0.011 | 103.43 | 0.000 | 1.12 | 1.16 |
| β SMB | 0.5203 | 0.018 | 29.45 | 0.000 | 0.49 | 0.55 |
| β HML | 0.1570 | 0.020 | 8.01 | 0.000 | 0.12 | 0.20 |
| High CSR Dummy | -0.0701 | 0.011 | -6.68 | 0.000 | -0.09 | -0.05 |
| Constant | 0.0886 | 0.004 | 21.2 | 0.000 | 0.08 | 0.10 |

## 5.2 Hypothesis 2 (H2)

The second hypothesis also uses the general KLD scores as a measure of CSP, but compares to performance of High CSR companies to companies that have a low score in the KLD social rankings (Low CSR).H2 therefore effectively tests the extremes when CSP is measured by simple KLD rating scores. H2a shows that when unadjusted for risk, high CSR companies do not significantly underperform low CSR companies statistically, but indicate that there is a difference in annual returns of 6%. This is an economically significant difference in annual returns.

| H2a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High CSR | 0.0102 | 0.0024 | 132 | 254 | -0.9070 | 0.1826 |
| Low CSR | 0.0162 | 0.0035 | 132 | | | |

H2b and H2c show that the higher level of returns for low CSR companies is accompanied by a much higher level of both traditional- and downside market risk. These results illustrate the wealth-protective effects of CSR and partly explain any financial outperformance by low CSR companies found under *a.*

| H2b | $B^{Market}$ | Robust Std. Error | t | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High CSR | 1.0224 | 0.027 | 37.67 | 0 | 0.97 | 1.08 |
| Low CSR | 1.1849 | 0.026 | 44.86 | 0 | 1.13 | 1.24 |

| H2c | $B^{BL}$ |
|---|---|
| High CSR | 1.057 |
| Low CSR | 1.204 |

Finally H2d shows that when adjusted for risk, returns on high CSR companies are significantly lower than returns on low CSR companies. This illustrates a large financial underperformance of over 9% for high CSR companies compared to low CSR companies. This is an increase of the plain difference in means that H2a found to be around 6%.

| H2d | Coef. | Robust Std. Error | t-value | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| $B^{Market}$ | 1.1158 | 0.019 | 57.74 | 0.000 | 1.08 | 1.15 |
| $\beta^{SMB}$ | 0.5028 | 0.031 | 16.11 | 0.000 | 0.44 | 0.56 |
| $\beta^{HML}$ | 0.1419 | 0.035 | 4.02 | 0.000 | 0.07 | 0.21 |
| High CSR Dummy | -0.0937 | 0.014 | -6.9 | 0.000 | -0.12 | -0.07 |
| Constant | 0.1084 | 0.009 | 11.83 | 0.000 | 0.09 | 0.13 |

**5.3 Hypothesis 3 (H3)**

The third hypothesis examines whether the distinction between frontier policies and civil foundation policies yields different results. In order to do so I test whether companies with high levels of frontier policies (High FP) and companies with high levels of civil foundation (High CF) perform different financially. Hypothesized is that the various stakeholders may have different or even opposing treatment for the two classes, which could lead to a difference in financial performance. The t-test shows that returns are not statistically different for the two portfolios. The difference in average returns of only 1% is compared to the other test hypotheses fairly low and therefore economically less significant.

| H3a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High FP | 0.0096 | 0.0028 | 132 | 261 | -0.1598 | 0.4366 |
| High CF | 0.0106 | 0.0025 | 132 | | | |

Sub-hypotheses *b* and *c* show surprisingly that companies with high scores on frontier policies are subject to more traditional market risk. The difference is however very small and even though not statistically verifiable, the confidence intervals indicate that there is no significant difference. More substantial is the difference when evaluating the more costly downside market risk.

| H3b | β Market | Robust Std. Error | t | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High FP | 1.0787 | 0.041 | 26.53 | 0 | 1.00 | 1.16 |
| High CF | 1.0585 | 0.022 | 47.43 | 0 | 1.01 | 1.10 |

| H3c | BL β |
|---|---|
| High FP | 1.122 |
| High CF | 1.085 |

When evaluating risk adjusted returns no statistically significant difference arises. The coefficient is negative at -0.0159, but as indicated by the p-value is not statistically different from zero. It appears to be so that stakeholders do not differentiate between high levels of CF or high levels of FP. If anything high FP companies underperform high CF, albeit insignificantly.

| H3d | Coef. | Robust Std. Error | t-value | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.0634 | 0.020 | 54.12 | 0.000 | 1.02 | 1.10 |
| β SMB | 0.2307 | 0.032 | 7.23 | 0.000 | 0.17 | 0.29 |
| β HML | 0.2185 | 0.036 | 6.08 | 0.000 | 0.15 | 0.29 |
| High FP Dummy | -0.0159 | 0.014 | -1.15 | 0.249 | -0.04 | 0.01 |
| Constant | 0.0484 | 0.009 | 5.34 | 0.000 | 0.03 | 0.07 |

**5.4 Hypothesis 4 (H4)**

The fourth hypothesis tests whether companies with a mid level score in civil foundation areas and a bad reputation (Mid CF - Bad Rep.) underperform the rest of the sample. This category also contains greenwashing companies that engage in moderate levels of civil foundation CSR, whilst having a bad reputation concerning CSR. Examining the a hypothesis no statistical significant difference in annual returns arises, but the coefficients indicate a difference of 2.5% in average annual returns. This difference can be deemed economically significant.

| H4a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| Mid CF - Bad Rep. | 0.0124 | 0.0025 | 132 | 259 | -0.3867 | 0.3496 |
| Other | 0.0149 | 0.0032 | 132 | | | |

The lower average returns for this sub-group can partly be explained by the lower market risk the companies are subject to. The difference in coefficients between the two groups is around 0.05 for both traditional and market volatility. However, the large standard error of the smaller Mid CF - Bad Rep. portfolio implies the two are not statistically different from one another. This is confirmed by the largely overlapping confidence intervals.

| H4b | β Market | Robust Std. Error | t-stat | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| Mid CF - Bad Rep | 1.0918 | 0.043 | 25.46 | 0 | 1.01 | 1.18 |
| Other | 1.1402 | 0.011 | 100.25 | 0 | 1.12 | 1.16 |

| H4c | BL β |
|---|---|
| Mid CF - Bad Rep. | 1.105 |
| Other | 1.167 |

When evaluating the results of H4d it becomes apparent that Mid CF – Bad Rep. companies significantly underperform the rest of the sample. On average companies belonging to this group show an underperformance of around 3% annually that can be attributed to the sub-group characteristic.

| H4d | Coef. | Robust Std. Error | t-value | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.1367 | 0.011 | 103.4 | 0.000 | 1.12 | 1.16 |
| β SMB | 0.5209 | 0.018 | 29.48 | 0.000 | 0.49 | 0.56 |
| β HML | 0.1585 | 0.020 | 8.09 | 0.000 | 0.12 | 0.20 |
| Mid CF- Bad Rep. Dummy | -0.0331 | 0.015 | -2.27 | 0.023 | -0.06 | 0.00 |
| Constant | 0.0831 | 0.004 | 20.7 | 0.000 | 0.08 | 0.09 |

**5.5 Hypothesis 5 (H5)**

The fifth hypothesis tests the difference between the Mid CF – Bad Rep. group and the companies that have low CF as well as low FP scores and a bad reputation (Low CSR – Bad Rep.). So how does the category containing greenwashing companies perform versus the one with socially undesirable companies. Examining the results of the t-test there is no statistically significant difference in mean returns between the two groups. The socially undesirables average returns are however slightly

higher than that off the greenwashing companies. This indicates that having no CSR or very low CSR produces slightly higher returns than having CF CSR, whilst having a bad reputation.

| H5a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| Mid CF - Bad Rep. | 0.0124 | 0.0025 | 132 | 233 | -0.1398 | 0.4445 |
| Low CSR - Bad Rep. | 0.0135 | 0.0053 | 132 | | | |

When examining the market risk associated with the two groups we see the higher returns of socially undesirables are easily explained as these companies have much higher market risk statistics. The absolute difference in traditional $\beta^{Market}$ is a whopping 0.42. The 95% confidence intervals don't overlap and therefore hint that this result would also be of some statistical significance. Furthermore, the difference in market risk between the two groups increases only further when considering the more costly downside market risk.

| H5b | β Market | Robust Std. Error | t | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| Mid CF - Bad Rep. | 1.0918 | 0.043 | 25.46 | 0 | 1.01 | 1.18 |
| Low CSR - Bad Rep. | 1.5109 | 0.066 | 22.74 | 0 | 1.38 | 1.64 |

| H5c | BL β |
|---|---|
| Mid CF - Bad Rep. | 1.105 |
| Low CSR - Bad Rep. | 1.541 |

When assessing the difference in financial performance as measured by risk-adjusted returns between the two groups, the one containing greenwashing companies shows a slight yet significant underperformance. The underperformance is only significant at the 10% confidence level which implies the result is statistically fairly weak.

| H5d | Coef. | Robust Std. Error | t-value | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.2685 | 0.038 | 33.71 | 0.000 | 1.19 | 1.34 |
| β SMB | 0.2659 | 0.057 | 4.67 | 0.000 | 0.15 | 0.38 |
| β HML | 0.2061 | 0.071 | 2.92 | 0.004 | 0.07 | 0.34 |
| Mid CF-Bad Rep. Dummy | -0.0503 | 0.027 | -1.89 | 0.059 | -0.10 | 0.00 |
| Constant | 0.0911 | 0.022 | 4.16 | 0.000 | 0.05 | 0.13 |

**5.6 Hypothesis 6 (H6)**

The sixth hypothesis tests the difference in financial performance between the group of social do-gooders or philanthropists and the rest of the sample. The philanthropist group is characterised by

high scores in FP areas and a good reputation on CSR (High FP – Good Rep.). Examining the results of the test for differences in means there is no significant under- or outperformance between the two sub-groups. However, the difference in annual averages of 4.3% can be marked as economically significant.

| H6a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 0.0107 | 0.0028 | 132 | 261 | -0.6413 | 0.2609 |
| Other | 0.0150 | 0.0031 | 132 | | | |

When evaluating the market risk of the two groups, the results of the *b* and *c* hypotheses indicate that the underperformance may partly be caused by a lower market risk factor for the High FP – Good Rep. group. The difference is however relatively small and the overlapping confidence intervals leave me hesitant to proclaim any real difference between the two. Also when evaluating downside risk, the difference between both β-figures is not substantial.

| H6b | β Market | Robust Std. Error | t | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 1.0901 | 0.049 | 22.15 | 0 | 0.99 | 1.19 |
| Other | 1.1388 | 0.011 | 101.04 | 0 | 1.12 | 1.16 |

| H6c | BL β |
|---|---|
| High FP - Good Rep. | 1.126 |
| Other | 1.164 |

Like expected when the difference in means is this large and the difference in market risk is only marginal, the evaluation of all factors leads to a statistically significant difference. The group of social outperformers show after controlling for other factors an average yearly underperformance of over 5% as compared to the rest of the sample.

| H6d | Coef. | Robust Std. Error | t-value | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.1366 | 0.011 | 103.4 | 0.000 | 1.12 | 1.16 |
| β SMB | 0.5211 | 0.018 | 29.49 | 0.000 | 0.49 | 0.56 |
| β HML | 0.1588 | 0.020 | 8.1 | 0.000 | 0.12 | 0.20 |
| High FP-Good Rep. Dummy | -0.0538 | 0.016 | -3.29 | 0.001 | -0.09 | -0.02 |
| Constant | 0.0834 | 0.004 | 20.97 | 0.000 | 0.08 | 0.09 |

**5.7 Hypothesis 7 (H7)**

H7 tests if companies that are true social outperformers, with a high frontier score and a good CSR reputation (High FP – Good Rep.), financially outperform companies with a good CSR reputation, yet without the high CSP (Good Rep.). When examining the test results from the $a$ test the High FP – Good Rep. group does not statistically under- or outperform the Good reputation group. The average returns for social outperformers are however on average 0.011 lower than of the good reputation group indicating that corporate philanthropy may be a costly practice.

| H7a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 0.0107 | 0.0028 | 132 | 262 | -0.1668 | 0.4338 |
| Good Rep. | 0.0118 | 0.0029 | 132 | | | |

When evaluating the market risk both groups are exposed to, there is no real difference in levels of market risk for both traditional and downside market risk. The completely overlapping confidence intervals support the low difference in coefficients. I therefore find no difference in market risk levels between social outperformers and companies that simply have a good CSP reputation.

| H7b | β Market | Robust Std. Error | t | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 1.0901 | 0.049 | 22.15 | 0 | 0.99 | 1.19 |
| Good Rep. | 1.1061 | 0.026 | 41.83 | 0 | 1.05 | 1.16 |

| H7c | BL β |
|---|---|
| High FP - Good Rep. | 1.126 |
| Good Rep. | 1.131 |

When examining risk adjusted returns, the High FP – Good Rep. group significantly underperforms the Good Rep. group. The coefficient is only significant at the 10% level and not at higher levels of statistical significance, but is significant nonetheless. So if anything, engaging in frontier policies when the corporate CSR reputation is already good is accompanied by lower financial performance.

| H7d | Coef. | Robust Std. Error | t-value | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.1019 | 0.023 | 47.25 | 0.000 | 1.06 | 1.15 |
| β SMB | 0.4442 | 0.036 | 12.45 | 0.000 | 0.37 | 0.51 |
| β HML | 0.1112 | 0.040 | 2.8 | 0.005 | 0.03 | 0.19 |
| High FP-Good Rep. Dummy | -0.0268 | 0.016 | -1.66 | 0.097 | -0.06 | 0.00 |
| Constant | 0.0631 | 0.011 | 5.51 | 0.000 | 0.04 | 0.09 |

**5.8 Hypothesis 8 (H8)**

The eighth hypothesis tests whether companies that are stuck-in-the-middle financially perform differently compared to social outperformers. The stuck-in-the-middle companies are characterised by mid level CF scores and a neutral reputation (Mid CF-Neutral Rep.). The first test hypothesis H8a indicates that there is neither a statistically significant nor an economically significant difference in average returns between the two sub-groups. The average returns are even slightly higher for stuck-in-the-middle companies.

| H8a | μ | σ | N | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 0.0107 | 0.0028 | 132 | 261 | -0.0794 | 0.4684 |
| Mid CF - Neutral Rep. | 0.0112 | 0.0031 | 132 | | | |

Zooming in on the other important factor of financial performance, market risk, there is no substantial difference in traditional market risk between the two groups. There is a difference however as the coefficient for stuck-in-the-middle companies is slightly higher, indicating some increased level of market risk. The difference is somewhat more substantial when evaluating downside market risk as the difference between both coefficients increases to almost 0.15. This indicates that stuck in-the-middle companies are subject to substantially more downside market risk than social outperformers.

| H8b | β Market | Robust Std. Error | t | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| High FP - Good Rep. | 1.0901 | 0.049 | 22.15 | 0 | 0.99 | 1.19 |
| Mid CF - Neutral Rep. | 1.1380 | 0.042 | 26.78 | 0 | 1.05 | 1.22 |

| H8c | BL β |
|---|---|
| High FP - Good Rep. | 1.126 |
| Mid CF - Neutral Rep. | 1.270 |

When interpreting the results of the full multiple regression there is no significant under- or outperformance between the two sub-groups. Based on the a to c tests this was to be expected as both effects were relatively small, causing significant under- or outperformance to be rather unlikely. In other words this implies stakeholders are unable to differentiate between social outperformers and those that are in the middle of the social spectrum.

| H8d | Coef. | Robust Std. Error | t-value | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.1204 | 0.032 | 34.65 | 0.000 | 1.06 | 1.18 |
| β SMB | 0.3855 | 0.046 | 8.44 | 0.000 | 0.30 | 0.48 |
| β HML | 0.0817 | 0.054 | 1.53 | 0.127 | -0.02 | 0.19 |
| High FP-Good Rep. Dummy | -0.0284 | 0.022 | -1.31 | 0.190 | -0.07 | 0.01 |
| Constant | 0.0521 | 0.013 | 3.88 | 0.000 | 0.03 | 0.08 |

## 5.9 Hypothesis 9 (H9)

Finally, the ninth hypothesis tests the other end of the extreme tested in H8. H9 tests whether stuck-in-the-middle companies underperform socially undesirables (Low CSR – Bad Rep.). Visible from *H9a* is no statistically significant difference in means. The difference in average annual returns of over 2% is however of some economical significance.

| H9a | μ | σ | n | df | t-stat | P(T<=t) |
|---|---|---|---|---|---|---|
| Low CSR - Bad Rep. | 0.0135 | 0.0053 | 132 | 246 | 0.2800 | 0.3898 |
| Mid CF - Neutral Rep. | 0.0112 | 0.0031 | 132 | | | |

The difference in means can very well be explained by the different exposures to traditional and downside market risk of both sub-groups. As the socially undesirables have very high exposure to market risk and the confidence intervals don't overlap I can assume there is a substantial difference in market risk between the two. The downside β figure mitigates this a little as for the Mid CF – Neutral Rep. group it is 0.12 higher compared to the traditional β figure and for the Low CSR – Bad Rep. group these figures are roughly the same.

| H9b | β Market | Robust Std. Error | t | P > \|t\| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| Low CSR - Bad Rep. | 1.5109 | 0.066 | 22.74 | 0 | 1.38 | 1.64 |
| Mid CF - Neutral Rep. | 1.1380 | 0.042 | 26.78 | 0 | 1.05 | 1.22 |

| H9c | BL β |
|---|---|
| Low CSR - Bad Rep. | 1.541 |
| Mid CF - Neutral Rep. | 1.270 |

When combining all elements in *H9d* we see that when controlling for market risk and the other factors, the Mid CF – Neutral Rep. companies significantly underperform the Low CSR – Bad Rep. companies. This indicates that doing no CSR at all and not spending any resources on CSR is possibly preferable to doing some Mid level CF with a neutral reputation.

| H9d | Coef. | Robust Std. Error | t-value | P > |t| | 95% Conf. Interval | |
|---|---|---|---|---|---|---|
| β Market | 1.2908 | 0.037 | 34.93 | 0.000 | 1.22 | 1.36 |
| β SMB | 0.3514 | 0.056 | 6.31 | 0.000 | 0.24 | 0.46 |
| β HML | 0.2334 | 0.071 | 3.31 | 0.001 | 0.10 | 0.37 |
| Mid CF - Neutr. Rep. Dummy | -0.0623 | 0.027 | -2.34 | 0.019 | -0.11 | -0.01 |
| Constant | 0.0847 | 0.022 | 3.87 | 0.000 | 0.04 | 0.13 |

**6. Discussion, Conclusion, Implications and Limitations**

**6.1 Discussion of results**

Before answering my research question and drawing my conclusions, I will first discuss my test results in more depth and in light of the existing theories on the CSP-CFP link. In my theoretical framework I have listed the various theories that explain the found results, I will now use these theories to evaluate my own test results.

To recall, the theories can be broadly defined in to two groups: those that suggest financial underperformance as a result of CSP and those that suggest financial outperformance as a result of CSP. A negative link may be found in the theory of Friedman (1970) that CSP is costly and will therefore be a waste of valuable company resources. Another negative link may be found in the agency problems that arise when company management can legitimize just about any action in the light of the interests of some stakeholder group.

Positive links are expected based upon the instrumental stakeholders theory that predicts various key stakeholders are more likely to maintain important ties with companies that have a higher CSP. Also, a positive link is predicted by the slack resources or good management theory. This theory expects companies that have time and resources available to focus on CSR to be well run and therefore likely outperforming other companies.

There are numerous academic papers that provide empirical evidence for any one of these theories. The slight majority as well as some important meta-analyses find a positive link between financial performance and corporate social responsibility. It is therefore the hypothesised relationships in my research build upon the instrumental stakeholder theory and expect a positive relationship. The results I encounter in my quantitative research do however not side with this view.

The first two hypotheses test the relationship between companies with a high CSR rating and respectively others in the sample (H1) and companies with a low CSR rating (H2). The results show that companies with a high CSR rating significantly underperform both other companies and companies with low CSR when measuring risk-adjusted returns. This is mainly caused by substantially lower average returns as high CSR rating companies are much less risky than their counterparts in H1 and H2 respectively.

The encountered negative relationship between CSP and CFP can be explained either negatively or positively. Both explanations are obviously contradictory, but they are both plausible and therefore worth discussing. A negative explanation would be based on the theory of Friedman and imply that

CSR is simply more costly than what it yields in indirect financial gains by the improved stakeholder ties. This would explain the financial underperformance of high CSP companies. A more positive explanation comes from the institutional stakeholder theory itself, since the theory predicts both negative and positive effects for stock returns. On the one hand it predicts that companies with high CSP have better ties with employees, customers and suppliers increasing profits and therefore stock returns. On the other hand it predicts that companies that have high CSP will have better ties with shareholders translating to a lower cost of capital. For equity cost of capital this would be observed as lower stock returns as a result of lower $\beta$ figures. The observed betas in both testing hypotheses 1 and 2, show that indeed for high CSR companies both traditional and downside betas appear to be substantially lower than for low CSR companies. This is in line with Sharfman and Fernando (2008) who find that firms with improved environmental risk management enjoy a lower cost of capital. Bollen (2007) shows that shareholders have multi-attribute utility curves and that the observed underperformance can be viewed as a lower boundary on the utility derived by shareholders from investing in socially responsible businesses, also pleading for more favourable financing options for companies with a high social rating.

Moving on, the third hypothesis tests whether any difference arises from the differentiation between frontier policies and civil foundation. Results show that there is no significant difference in financial performance between the two, indicating that stakeholders are unable to differentiate between CF and FP. Average returns are slightly higher (1% annually) for high CF than for high FP scoring companies, but also this difference is statistically insignificant. The difference has some economical significance and explained negatively indicates that FP is more costly than CF to undertake. This could be expected as less direct stakeholders are genuinely involved. Explained positively, the difference indicates that stakeholders recognise the value of FP over CF and therefore accept lower levels of stock returns. The found effect is of no statistical significance but if anything the negative link between CSP and CFP is amplified when measuring CSP with frontier policies.

The fourth hypothesis tests whether companies that have medium levels of CF and a bad reputation, underperform the rest of the sample. This is the category that among others contains companies that engage in greenwashing. Although not all companies in this category actively choose to greenwash, the fact that companies in this category engage only mildly in civil foundation policies implies that their social policy is unlikely to be driven by corporate philanthropy. If it were, they would most likely engage in frontier policies and the resulting CSR reputation would not be bad. Therefore, even though maybe not actively greenwashing, the combination of mid CF and bad reputation is a hypocritical one at the least. After controlling for several control variables, I find that this hypocrisy indeed leads to underperformance of the rest of the sample. However, the hypothesis was built on

the idea that social outperformance would lead to financial outperformance, so before drawing conclusions from this the results from H4 must be regarded in light of the results from H1 to H3. H1 to H3 show a clear negative link between increased levels of CSR and financial performance. Therefore the underperformance of Mid CF-Bad Rep. companies is in line with the expected underperformance as a result of increased levels of CSR. It appears to be so that stakeholders reward greenwashing similar to a sincere increase in CSR. The underperformance exhibited by greenwashing companies is lower than the underperformance exhibited by social outperformers. Explained negatively this implies that greenwashing is less costly than social outperformance. Explained positively this implies that greenwashing is rewarded with a reduced cost of capital. Both options indicate that stakeholders are unable to differentiate between hypocrisy or greenwashing and any other increase in CSR levels.

This effect is also clear when evaluating the results of H5 that compares Mid CF-Bad Rep. companies to socially undesirables (low CSR and bad reputation). This hypothesis tests what mid level CF brings compared to low CSR i.e. what does it yield a company to do some mid level of CSR if it already has a bad reputation. The tests show a large reduction in market risk, but also significant financial underperformance as measured by risk-adjusted returns as a result of increasing CSP to mid level CF. This indicates that either shareholders reward the company for the mid level CF with a lower cost of capital, even when reputation is bad, or that mid level CF is costlier than no or low CSR causing lower returns. In summary, I find no support for the research of Yoon et al. (2006) and Godfrey (2005) suggesting  the effect that CSR has on stakeholders to be dependent on the sincerity of company motives. My results show that greenwashing and contradictory CSP is treated similar to other increases in CSP and is regardless of any insincerity of company motives rewarded by stakeholders. This indicates that stakeholders are either unable to identify company motives or have no regard for company motives.

Furthermore the sixth and seventh hypothesis show that social outperformers (High FP – Good Reputation) are significant financial underperformers as measured by risk-adjusted stock returns, both compared to the rest of the sample as well as to other good reputation companies. This indicates either a large reduction in cost of capital for their excellent social performance or a large drop in returns because of undertaking costly FP. However, the rather small differences in beta figures suggest that a reduction in the equity cost of capital is less likely as I would expect to find lower betas as well. This implies that real social outperformance is costly and stakeholders do not appreciate company efforts to the level of expenses inquired. The link between CSP and CFP can be said to be significantly negative.

The last two hypotheses are interesting to evaluate whether a company would be better off picking an extreme CSR strategy i.e. social outperformance and socially undesirable or do a bit of both and find themselves in the middle of the CSP spectrum. I refer to this position as stuck-in-the-middle which is characterised by mid level CF and a neutral reputation. My tests uncover no statistically significant difference in financial performance between social outperformers and those that are socially stuck-in-the-middle .Test results from *H9* show significant financial underperformance when stuck-in-the-middles are compared to the socially undesirables. The betas found in *H9b* point out that stuck-in-the-middle is rewarded with a significantly lower cost of capital versus socially undesirables. However, *H9d* shows that when adjusted for risk, stuck-in-the-middle companies still exhibit significant financial underperformance. Whether stuck in the middle is a good strategy depends completely on the interpretation. If I take the positive explanation, apparently stuck-in-the-middle yields similar reduction in cost of capital to social outperformance at a lower cost and is therefore a pretty good strategy. If I follow the negative explanation, stuck-in-the-middle is a bad strategy as it apparently is roughly equally costly as social outperformance, which when costs are equal is preferable because of network externalities

## 6.2 Conclusion

The aim of this study has been to find a relationship between corporate social performance (CSP_ and corporate financial performance (CFP). My interest in the puzzle that is the CSP-CFP link started already in my bachelor thesis project and I was able to research this puzzle more in depth for this thesis. With my research I have set out on a search to answer my research question:

Does there exist a relationship between Corporate Social Performance and Financial Performance when both are measured effectively?

In order to do so, I will use the conclusions I can draw from my empirical analysis. Based on the instrumental stakeholders theory, I expected and hypothesised that the sign of the relationship between CSP and CFP would be positive. I expected effects to be amplified in some extreme cases such as with social do-gooders and to be mitigated in others as for example with greenwashing. My results do not side with this view and clearly indicate a negative relationship between CSP and CFP. However, I still observed some of the hypothesized effects, such as amplification in the extremes, albeit with an opposing sign. Therefore my results do not completely break with existing theory. Also, forming an expectation on the sign was necessary for testable hypotheses, but there is a reason why I have refrained to do so in forming my research question. The large body of academic evidence for both positive and negative CSP-CFP links result in that test outcomes showing a negative CSP-CFP link are unanticipated, but not unexpected.

The first conclusion I can draw from my research is that there is a negative link between levels of CSP and financial performance. The link appears to be positive when considering market risk both measured by traditional β's and downside β's and is negative when considering returns. However, when all factors are considered the risk adjusted returns exhibit an unambiguously negative relation with CSP. This essentially answers my research question: there exists a relationship between CSP and FP when both are measured effectively.

This isn't however the most important result of this study. The most important find is how the various stakeholders involved with a company respond to CSP. So are stakeholders able to differentiate between FP, CF and simple CSR ratings? I expected traditional measures of CSR to not be able to establish an unambiguous relationship between CSP and CFP, because it would hold no information on how stakeholders perceive CSP. I hypothesized that the many contradictory academic papers were the result of the aggregation of contradictory effects and measurement errors of either financial performance or CSP. In light of this, I have several important findings.

The first one is that stakeholders are unable to differentiate between actual social performance and social ratings in the first place. This is supported by my find that stakeholders do not treat high FP companies different from high CF companies. This is remarkable as it indicates that stakeholders are unable to differentiate between CF policies, which are often more tightly linked to direct stakeholders such as employees, and FP policies, which tend to have a more philanthropic nature. Add to this the results of H1 and H2 and it is clear that stakeholders are unable to differentiate between high ratings, high CF and high FP. The only plausible explanation for this is that ratings are more important than I expected. It appears to be so that stakeholders assess CSP quite superficially and social ratings are very important in the social perception, even more so than actual social performance is. This is in correspondence with Queré, Nouyrigat & Baker (2015) that emphasize the importance of social ratings for European listed companies.

Another finding that supports the theory that stakeholders are unable to identify or have no regard for actual social performance is that greenwashing, or at least contradictory CSP, is rewarded similar to an increase in CSP when the company reputation is good or neutral. That is, stakeholders are indifferent whether a company motivation is sincere or insincere and judge an increase in CSP independent of corporate reputation.

The second important finding is that using more in depth measures of financial performance is vital in capturing the full effects of CSP on financial performance. I have measured financial performance by measuring both returns and risk on a stand-alone basis as well as risk-adjusted returns. In addition I have used measures of downside-risk to support results of conventional risk

measures. The use of the various measures allows me to capture all effects and only using one would attribute too much difference to the CSP characteristic. The use of risk as a standalone variable adds valuable information as the wealth-protective effects of CSP are only visible when risk is assessed on a standalone basis. This information can be used in optimizing social investing strategies.

So in conclusion, my study has yielded an unambiguous answer on my research question as there exists a clear negative link between CSP and CFP. Furthermore, the study provided valuable insights in stakeholder perception of CSP. Stakeholders are unable or unwilling to differentiate between actual social performance and reported CSR. This implies that social ratings are of vital importance for stakeholder perceptions even more important than actual social performance. Also measurement of financial performance should be done by using both risk-adjusted returns as well as risk and return characteristics on a stand-alone basis.

**6.3 Implications and suggestions for further research**

An important conclusion is that differentiating between actual social performance and social ratings does not yield all too different results for the sign of the CSP-CFP link. I expected ambiguous results in the body of prior academic research, to be as result of CSP measures not measuring CSP or financial performance measures being too superficial. My results indicate that this is not the source of the varying research conclusions, as stakeholders are unable or unwilling to differentiate between actual social performance and social rating scores. The implication is that the debate is still out on the reason why research outcomes on this subject have such contradictory results.

Furthermore, illustrated in the above discussion of the results, there are multiple explanations for the observed negative relationship between CSP and CFP. Implications of my study are largely dependent on whether the true explanation of this relationship is positive or negative. When companies are rewarded with a lower cost of equity capital for their CSP levels, implications are quite the opposite from when underperformance is a result of a simply costly expense. I will discuss the implications under both scenarios.

In the first scenario companies are rewarded a lower cost of equity capital for their CSP levels. In this case companies should increase their CSR in order to increase their social ratings. This should be done with the most visible and least costly (CF) type of CSR policies, as shareholders tend to be indifferent between the two. For companies that have a bad reputation, even a slight increase in visible CF policies can reduce the cost of capital significantly. This proves greenwashing to be a very rewarding strategy.

In the second scenario CSP is simply more costly than beneficial to the company as company assets are spent on activities that do not directly benefit the company. If this is the true reason for CSP related underperformance, CSP should be as low as possible. Greenwashing is in this scenario a terrible strategy as it reduces financial performance without any beneficial effects.

Further research should focus on determining the true explanation for the found phenomena. This could be done by for example examining the cost of debt capital to find certain parallels in financing benefits among companies with high social rankings. Again other research could include accounting measures next to market measures to determine whether financial underperformance is present in accounting measures. If accounting measures show similar underperformance a theory of cost of capital reduction is less likely.

**6.4 Limitations**

A similarity between the two scenarios of implications is that there is no difference in financial performance for a company to engage in frontier policies and to engage in civil foundation policies. This study does not imply that there are no arguments for engaging in frontier policies, but merely shows that, for large U.S. corporates , stakeholders are indifferent between frontier and civil foundation policies. This is also the first limitation of this study. As I wanted to perform a quantitative study on financial performance, I needed vast quantities of data. The only way to achieve this was using listed companies. The issue with the majority of the currently listed companies is often however, that doing good is optional and not essential to the business model. I have corresponded with U.S institute B-lab that certifies certain companies as B-Corporations or B-corps. The certification follows very strict criteria and focuses on businesses that use "Business as a force for good". Although having graded over 1,800 companies, only a handful of these was listed, which makes quantitative research close to impossible on their label. It also indicates that possibly smaller companies are the real social outperformers and stakeholders may treat social performance at the core of business very differently from optional CSP. A customer may for example decide to buy products at a regional socially responsible store and not at Wal-Mart because of social considerations, but is less likely to shift from Cost-Co to Wal-Mart purely on social considerations. This difference is emphasized by the difference in governance forms. As the LLC (Dutch B.V.) allows for more shareholder activism compared to the public listing, with many unknown shareholders and often large institutional shareholders like mutual and pension funds that have their own mandates. It may therefore well be that listed companies isn't the sample to draw conclusive evidence from when it comes to social performance.

Next to finding the true explanation of the found effects, further research should therefore focus on scaling down to smaller samples of unlisted social outperformers. As labels like B-corporation develop, these databases are created automatically and research can (with access to financial information of these private companies) concentrate on how stakeholders perceive CSP as the main business objective and how this translates to financial performance. Other research can concentrate on scaling down of frontier policies and pick out a few individual frontier policies for an event study that are truly revolutionary.

Another limitation is the timeframe. As the KLD data was only available until 2013, my sample period is already "recent history". Stakeholder perception is ever changing and 2003 is a long time ago, also culturally. The link between CSR and CSP should be examined again in later periods, since changing stakeholder perceptions and societies will change the link between CSP and CFP.

To conclude, there are still many steps to take in fully uncovering the puzzle that is the CSP-CFP link. My study finds a negative relationship between corporate social responsibility and financial performance for listed U.S. companies from 2003 to 2013. Furthermore, I find that stakeholders are unable or unwilling to differentiate between actual corporate social performance and social ratings in the first place.

**References**

Anderson, J., & Frankle, A. (1980). Voluntary Social Reporting: An Iso-Beta Portfolio Analysis. *The Accounting Review , 55* (3).

Ang, S., & Wight, A. (2009). Building intangible resources: The stickiness of reputation. *Corporate Reputation Review , 12* (1), 21-32.

Aupperle, K., & Van Pham, D. (1989). An Expanded Investigation in the Relation-ship of Corporate Social Responsibility and Financial Performance. *Employee Respon-sibilities and Rights Journal* , 263-276.

Aupperle, K., Caroll, A., & Hatfield, J. (1985). An Empirical Examination of the Relationship Between Corporate Social Responsibility and Profitability. *The Academy of Management Journal , 28* (2), 446-463.

Bae, J., & Cameron, G. T. (2006). Conditioning effect of prior reputation on perception of corporate giving. *Public Relations Review , 32* (2), 144-150.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance , 29*, 1751-1767.

Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics , 2*, 95-121.

Bawa, V. S., & Lindenberg, E. B. (1977). Capital Market Equilibrium in a Mean-Lower Partial Moment Framework. *Journal of Financial Economics , 5*, 189-200.

Bollen, N. (2007). Mutual Find attributes and investor behaviour. *Journal of Financial and Quantative Analysis , 42* (3), 683-708.

Bragdon, J., & Marlin, J. (1972). Is pollution profitable? *Risk Management , 19* (4).

Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate Social Performance and Stock Returns: UK Evidence from Disaggregate Measures. *Financial Management , 35* (3), 97-116.

Brammer, S., Millington, A., & Rayton, B. (2007). The contribution of corporate social responsibility to organizational commitment. *International Journal of Human Resource Management , 18*, 1701-19.

Carroll, A. (1979). A Three Dimensional Conceptual Model of Corporate Performance. *Academy of Management , 4* (4), 497-505.

Carroll, C. E. (2010). *Corporate Reputation and the News Media: Agenda-setting within business news coverage in developed, emerging, and frontier markets.* Routledge.

CERES & Sustainalytics . (2013). *The Road to 2020: Corporate Progress on the Ceres Roadmap for Sustainability.*

CERES. (2013). *CERES Sustainability Report 2013.*

CERES. (2012). *Hydraulic Fracturing & Water Stress: Water Demand by the Numbers.*

Cox, P., Brammer, S., & Millington, A. (2004). An empirical examination of institutional investor preferences for corporate social performance. *Journal of Business Ethics , 52*, 27-43.

Emmelhainz, M., & Adams, R. (1999). The apparel industry response to "sweatshop" concerns: a review and analysis of codes of conduct. *Journal of Supply Chain Management , 35* (2), 51-57.

Fama, E. F., & French, K. R. (1992). The Cross-Section of Expected Returns. *Journal of Finance , 2*, 427-465.

Fombrun, C. J., Gardberg, N. A., & Barnett, M. (2000). Opportunity Platforms and Safety Nets: Corporate Citizenship and Reputational Risk. *Business and Society Review , 105* (1), 85-106.

Fombrun, C., & Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *Academy of Management , 33* (2), 233-258.

Frederick, W. (1994). From CSR1 to CSR2. *Business & Society , 33*, 150-164.

Freeman, R. (1994). *Strategic management: A stakeholder approach.* Boston, MA: Pitman Press.

Friedman, M. (1970, September 13). The social responsibility of business is to increase its profits. *New York Times* .

Frynas, J. (2005). The false developmental promise of Corporate Social Responsibility: evidence from multinational oil companies. *International Affairs , 81* (3), 581-598.

Geczy, C., Stambaugh, R., & Levin, D. (2003). Investing in Socially Responsible Mutual Funds. *Working Paper* .

Godfrey, P. (2005). The relationship between coporate philantropy and shareholder wealth: A riks management perspective. *Academy of Management Review , 30*, 777-798.

Godfrey, P. (2005). The Relationship Between Corporate Philanthropy And Shareholder Wealth: A Risk Management Perspective. *Academy of Management Review , 30* (4), 777-798.

Griffin, J., & Mahon, J. (1997). The corporate social performance and financial performance debate: twenty-five years of incomparable research. *Business & Society , 36* (1), 5-31.

Guenster, N., Bauer, R., Derwall, J., & Koedijk, K. (2010). The Economic Value of Corporate Eco-Efficiency. *European Financial Management , 17* (4), 679-704.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing Well While Doing Good The Investment Performance Of Socially Responsible Mutual Funds. *Financial Analysts Journal , 49* (6), 62-66.

Hillman, A., & Keim, G. (2001). Shareholder value, stakeholder management, and social issues: what's the bottom line? *Strategic Management , 22*, 125-39.

Jensen, M. (2001). Value Maximization, Stakeholder Theory, And The Corporate Objective Function. *Journal of Applied Corporate Finance , 22* (1), 32-42.

Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics , 4* (3), 305-360.

Jones, T. (1995). Instrumental stakeholder theory: A synthesis of ethics and economics. *Academy of Management review , 20*, 404-437.

Khanna, M., & Damon, L. (1999). EPA's Voluntary 33/50 Program: Impact on Toxic Releases and. *Journal of Environmental Economics and Management , 37* (1), 1-25.

Lyon, L. J., & Cameron, G. T. (2004). A Relational Approach Examining the Interplay of Prior Reputation and Immediate Response to a Crisis. *Journal of Public Relations Research , 16* (3), 213-241.

Maloni, M., & Brown, M. (2006). Corporate social responsibility in the supply chain: An application in the food industry. *Journal of Business Ethics , 68* (1), 35-52.

Margolis, J., & Walsh, J. (2003). Misery Loves Companies: Rethinking Social Initiatives by Business. *Administrative Science Quarterly , 48*, 268-305.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance , 7* (1), 77-91.

Martin, R. (2002). The virtue matrix. Calculating the return on corporate responsibility. *Harvard Business Review , 80* (3), 68-75.

Mattingly, J. (2015). Corporate Social Performance: A Review of Empirical Research Examining the Corporation–Society Relationship Using Kinder, Lydenberg, Domini Social Ratings Data. *Business & Society* , 1-44.

Mcguire, J., Sundgren, A., & Schneeweiss, T. (1988). Corporate Social Responsibility and Firm Financial Performance. *The Academy of Management Journal , 31* (4), 854-872.

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal , 21* (5), 603-609.

Mcwilliams, A., & Siegel, D. (2001). Corporate Social Responsibility: A Theory of the Firm Perspective. *Academy of Management , 26* (1), 117-127.

Oikonomou, I., Brooks, C., & Brammer, S. (2012). The Impact of Corporate Social Performance on Financial Risk and Utility: A Longitudinal Analysis. *Financial Management* , 483-515.

Orlitzky, M., Schmidt, F., & Rynes, S. (2003). Corporate social and financial performance: A meta-analysis. *Organization studies , 24*, 403-411.

Preston, L., & O'Bannon, D. (1997). The Corporate Social-Financial Performance Relationship-A Typology and Analysis. *Business & Society , 36* (4), 419-429.

Quére, B. P., Nouyrigat, G., & Baker, C. R. (2015). A Bi-Directional Examination of the Relationship Between Corporate Social Responsibility Ratings and Company Financial Performance in the European Context. *Journal of Business Ethics* .

Renneboog, L., Ter Horst, J., & Zhang, C. (2011). Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. *Journal of Financial Intermediation , 20* (4), 562-588.

Rennenboog, L., Ter Horst, R., & Zhang, C. (2008). Socially responsible investments : institutional aspects, performance, and investor behavior. *Journal of Banking and Finance , 32* (9), 1723-1742.

Ruf, B., Muralidhar, K., & Paul, K. (1998). The development of a systematic, aggregate measure of corporate social performance. *Journal of Management , 24* (1), 119-133.

Salama, A., Anderson, K., & Toms, J. (2009). Does community and environmental responsibility affect firm risk? Evidence from UK panel data 1994–2006. *Business Ethics , 20* (4), 192-204.

Sen, S., & Bhattacharya, C. (2001). Does doing good always lead to doing better? Consumer reactions to corporate social responsibility. *Journal of Marketing Research , 38*, 225-43.

Shahzad, A., & Sharfman, M. P. (2015). Corporate Social Performance and Financial Performance Sample-Selection Issues. *Business & Society* , 1-30.

Sharfman, M., & Fernando, C. (2008). Environmental Risk Managmement And The Cost of Capital. *Strategic Management Journal , 29*, 569-592.

The Economist. (2015, April 22). In the bin, recycling in America. *The Economist* .

Tirole, J. (2001). Corporate Governance. *Econometrica , 69* (1), 1-35.

Ullman, A. (1985). Data Is Search of a Theory: A Critical Examination of the Relation-ships Among Social Performance, Social Disclosure, and Economic Performance. *Academy of Management Review , 10*, 540-577.

Vidovic, M., & Khanna, M. (2007). an voluntary pollution prevention programs fulfill their promises? Further evidence from the EPA's 33/50 program. *Journal of Environmental Economics and Management , 53*, 180-195.

Waddock, S., & Graves, S. (1997). The corporate social performance-financial perfomance link. *Strategic Management Journal , 18* (4), 303-319.

Wood, D. (1991). Corporate Social Responsibility Revisited. *Academy of Management , 16* (4), 691-718.

Yoon, Y., Zeynep, G., & Schwarz, N. (2006). The Effect of Corporate Social Responsibility (CSR) Activities on Companies With Bad Reputations. *Journal of Consumer Psychology , 16* (4), 377-390.

Zweigenhaft, R. L., & Domhoff, G. (2011). *The New CEOs: Women, African American, Latino, and Asian American Leaders of Fortune 500 Companies.* Lanham: Rowan & Littlefield.