Giving CSR a run for its money *

Sean Nordqvist[1] & Daniel Raber[2]

Bachelor Thesis in Finance

Stockholm School of Economics

May 16, 2016

Abstract

This paper studies the link between Corporate Social Performance (CSP) and Corporate Financial Performance (CFP). We find no indication of an existing relationship between CSP and accounting/market-based CFP when using an aggregated CSP score, in support of neutrality theory. When expanding our analysis by disaggregating CSP into environmental, social and corporate governance (ESG) domains, we observe that results differ depending on what CSR rater is used to measure CSP. We compare two major CSR raters, MSCI KLD and Thomson Reuters ASSET4, and show that there is moderate convergent validity between the two, which could be a major reason why results in this field have over the past 35 years been heterogeneous. Building on legitimacy theory, our results suggest that companies in polluting industries have more to gain financially from adopting environmental practices than do companies in non-polluting industries. After using numerous econometric techniques to partially account for potential endogeneity, our results do not change significantly.

Keywords: Corporate Social Responsibility, Corporate Financial Performance, Endogeneity, Convergent Validity, Legitimacy Theory

---

* We would like to thank our tutor Daniel Metzger for his insightful comments and support during the process of writing this thesis. We also extend our gratitude to Giovanni Battista Derchi (University of Lausanne) for his helpful guidance. Lastly, we thank MSCI for giving us access to the MSCI ESG KLD STATS database. All errors are our own.

[1] 23003@student.hhs.se

[2] 23054@student.hhs.se

Table of contents

List of abbreviations

CFP	Corporate financial performance
CSP*	Corporate social performance
CSR	Corporate social responsibility
ESG	Environmental, Social, Governance
FE	Fixed effects
IV	Instrumental Variable
OVB	Omitted-variable bias
R&D	Research and development
RE	Random effects
ROA	Return on assets
ROE	Return on equity
SUR	Seemingly unrelated regressions

*CSP is used interchangeably with CSR score and CSR rating in this paper

1 Introduction

1.1 CSP and CFP - a complicated relationship

Since the release of Howard Bowen's book "Social Responsibilities of the Businessman" in 1953, corporate social responsibility has grown to become an increasingly integral part of today's business world. Corporate social responsibility, or CSR, is a self-governing mechanism that firms adopt in order to ensure that their business is conducted in a responsible manner and has a positive impact on society. CSR is measured based on environmental, social and governance (ESG) aspects of a firm's business activities. Today, many investors screen investment decisions based on the extent to which a firm is socially responsible, and firms that do not take CSR seriously are often subject to public criticism and scrutiny (TIME, 2012).

The decision for a firm to adopt CSR is not entirely clear. Opponents of CSR argue, in line with Friedman (1970) and shareholder theory, that the managerial fiduciary duty is to maximize shareholder value rather than focusing on meeting the demands of other stakeholders. Proponents, however, claim that a stakeholder-theory type of view created by being socially responsible will lead to advantages that can ultimately translate into improved financial performance for a firm. Thus, much of the CSR related research has centered on trying to establish a link between a firm's corporate social performance (CSP) and its corporate financial performance (CFP). The findings of these studies are heterogeneous suggesting positive, negative, neutral and even bidirectional relationships (e.g. Vance, 1975; McGuire et al., 1988; Waddock and Graves, 1997; Orlitzky et al., 2003). According to Ullmann (1985), the mixed findings are mainly due to three reasons: (a) a lack in theory, (b) inappropriate definition of key terms and (c) deficiencies in empirical databases currently available. Building on this reasoning by Ullmann (1985), this study re-examines the link between CSP and CFP but during recent years. In order to address any potential deficiencies in empirical databases currently available, this study unlike any other uses two different databases in order to compare the results pertaining to the CSP and CFP relationship. Moreover, this study examines industry specific mechanisms of the CSP and CFP link.

1.2 Thesis structure and summary

The first part of this study analyzes the link between CSP and CFP for 400 listed companies in the US. As endogeneity is a central issue when examining this link, this issue is a main focus in this paper. We find that it is not uncommon for studies in this field to be on the weak end of accounting for endogeneity, which could be a contributing factor to the heterogeneous results.

In order to measure CSP, two empirical databases are used: KLD[3] and ASSET4[4]. The main focus in the study lies on KLD, but ASSET4 as a proxy for CSP is used in order to see if results vary depending on what empirical database is used. As proxies for CFP, a market-based performance measure, Tobin's Q, and an accounting-based performance measure, ROE, are used. When KLD is used as a proxy for CSP, no relationship is found between CSP and CFP. This can be explained using neutrality theory, which stipulates that due to the high degree of intermediary and moderating variables there is at best only a random link between CSP and CFP. When ASSET4 is used as a proxy for CSP, the results are somewhat different. When Tobin's Q is used to measure CFP, a weak negative link between the governance dimension and CFP is found. When using ROE a positive relationship between the environmental dimension and CFP is found, as well as a negative relationship between the social dimension and CFP. The differences in results obtained when using KLD or ASSET4 is likely to be due to low degrees of theorization (i.e. what the empirical databases actually measure) and commensurability (i.e. how CSP is measured) (Chatterji, et al., 2015).

In order to further examine the differences regarding the CSP-CFP relationship that is found when KLD or ASSET4 is used, we look at the relationship between KLD and ASSET4 across years by means of their convergent validity. A moderate degree of convergence between the two empirical databases is observed, thus strengthening the argument that any differences are due to theorization and commensurability. An important implication of the analysis between KLD and ASSET4 is that future research studying the CSP-CFP link should preferably use several empirical databases for CSP in order to validate that the nature of the CSP-CFP relationship

[3] KLD is now part of the MSCI KLD ESG STATS database. We refer to it as KLD throughout. See data disclaimer in **section 12**.
[4] ASSET4 is an ESG product by Thomson Reuters.

that is found is not limited to a certain database. Furthermore, the results from comparing KLD and ASSET4 should be kept in mind when comparing findings from previous studies.

To assess any industry specific mechanisms behind the CSP and CFP relationship, the sample used in this study is divided into two groups: Group (1) are firms that belong to industries that can be classified as high-polluting, and group (2) are firms that belong to industries that are low-polluting. This study examines if firms in high-polluting industries have more to gain from being environmentally responsible (i.e. strong performance in the environmental dimension of ESG) in CFP than do firms in non-polluting industries. We find that the environmental dimension for firms in group (2) is negatively correlated with CFP. Using legitimacy theory, this could be explained by the fact that firms in group (2) do not have a "bridge to gap" between their actions and social values and norms, as they do not pollute.

When as a robustness check the measure of CSP is lagged with more than one year which is the case in the main model, the results suggest that firms in group (1) have more to gain in CFP from strong performance in the environmental dimension than group (2). This too can be explained by legitimacy theory in that firms in group (1) are likely to be perceived as more "sinful" due to the pollution and thus CSR activities will demonstrate that these firms are not actually "that bad". This, e.g. improved firm reputation, is then translated into CFP according to the results. These industry specific findings are useful as they provide with a better understanding of the mechanics underlying the CSP-CFP relationship. They also increase the understanding of why firms in different industries voluntarily choose to adopt CSR related practices, as can be explained using legitimacy theory.

To summarize, this study provides an extensive analysis of the CSP-CFP relationship with an emphasis on accounting for endogeneity. Contrary to prior studies in this field, this study addresses a main issue relating to the fact that empirical databases measuring CSP might differ from each other. Furthermore, the industry specific findings provide us with a better understanding of how the mechanics behind the CSP-CFP might differ depending on industry type. It is safe to say that we give CSR a good run for its money.

2 Previous Literature

Among studies attempting to establish a relationship between CSP and CFP, the findings are certainly mixed. Some empirical studies indicate either a positive, negative, bi-directional or non-existent linkage, while other academic research conclude that the standard methodologies applied by a vast majority of studies are econometrically flawed (Garcia-Castro, et al., 2010). In this section, previous literature examining the link between CSP and CFP will be presented. In **section 2.1**, research concluding a positive and/or bidirectional relationship between CSP and CFP will be discussed. In **section 2.2**, literature relating to a negative relationship between CSP and CFP will be presented. In **section 2.3**, empirical studies concluding none or an ambiguous relationship between CSP and CFP will be provided. In **section 2.4**, literature relating to the validity of different CSR ratings will be discussed. Literature relating to differences in CSR practices depending on industry type will be presented in **section 2.5**. Finally, contribution to existing literature will be discussed in **section 2.6**.

2.1 Positive and bidirectional relationships between CSP-CFP

In line with the findings of Cochran and Wood (1984), most studies focus on either accounting-based returns or market-based returns as proxies for CFP when examining the CSP-CFP relationship. An early study by Moskowitz (1972) examines the short-term share performance of 14 listed firms in the US that were perceived as socially responsible. In this study, Moskowitz finds that share returns of these selected socially responsible firms outperform several major indices. Although the study has received criticism due to the seemingly subjective way of measuring social responsibility as well as the lack of adjustment for systematic risk, the study at the time was an important contribution and stepping stone in the research field of CSR as it showed the value of screening portfolios (Cochran and Wood, 1984; Ullman, 1985).

In a comprehensive study by Waddock and Graves (1997), the empirical linkages between CSP and CFP are examined using a constructed CSR index based on the KLD index structure for a sample of 469 companies in the S&P 500. The study finds a positive relationship between

CSP and accounting-based CFP as measured by ROA, ROE and return on sales – i.e. indicating that a company can "do well by doing a good". Plausible explanations for the positive relationship includes improved investor relations, company reputation and employee moral by being, or at least appearing to be, socially responsible. Interestingly, the study also finds that CSR depends on financial performance indicating that companies can also "do good by doing well".

Orlitzky et al. (2003) examine the relationship between CSP and CFP by conducting a meta-analysis of 52 previous studies. The study concludes a positive relationship between CSP and CFP, and shows that the strength of the positivity varies depending on if market-based or accounting-based performance measures are used. The correlation between CSR and financial performance is found to be higher with accounting-based measures. Similar to the abovementioned findings by Waddock and Graves (1997), a bidirectional relationship between CSR and financial performance is found in the study.

In a study by Schreck (2011), where CSR-scores obtained from oekom research AG are tested against the market-based financial performance indicator Tobin's Q, no bidirectional relationship as found by Waddock and Graves (1997) and Orlitzky et al. (2003) can be established. The study, however, concludes a positive relationship between the corporate governance and environmental management components of CSP and Tobin's Q.

On the subject of bi-directionality, Granger (1969) proposes a test which can be used to assess whether one future time series has any predictive effect on a lagged time series. If a lagged time series CSP is significantly correlated with current time series CFP, but lagged values of CFP show no correlation current values of CSP, then CSP 'Granger causes' CFP. The Granger causality test has been used in CSP-CFP literature with mixed results. Shreck (2011) applies this methodology and finds no evidence that CSP would Granger cause CFP. The test has been criticized by econometricians as a case of *post hoc fallacy*[5] (Diebold, 2000). Moreover, it is not directly applicable to longitudal/panel data, which is the prevalent construction in a majority of studies in this field.

[5] States that the order of events (i.e. X occurred before Y, therefore X must cause Y) is not a sole determinant of causality

Cochran and Wood (1984) find a weak positive link between CSP and CFP using a total sample of 61 firms spread over the time periods 1970-1974 and 1975-1979. The study uses accounting returns as performance measures and at the time of the study expanded CSR research by controlling for asset age, a control variable that the authors of the study find to be highly correlated with CSP. The weak positive link between CSP and CFP remains after testing for asset age. In their study, Cochran and Wood argue that the results and conclusions regarding the relationship between CSP and CFP in previous similar research that ignores controlling for asset age should be second-guessed. Furthermore, the study also includes an industry-specific control group.

In 1988, McGuire et al. find an overall positive relationship between CSP and CFP. The study, often quoted by CSR researchers due to its extensive nature, measures CSR using data from a survey created by Fortune and measures both how CSP predicts CFP and how CFP predicts CSP. An overall positive relationship between CSP and CFP is found using both market-based and accounting-based measures, such as Alpha and ROA. McGuire et al. also conclude that accounting-based to a higher degree seems to correlate with CSP than do market-based measures and that prior CFP appears to be a superior predictor of CSP rather than subsequent CFP.

2.2 Negative relationship between CSP and CFP

Central to literature supporting a negative relationship between CSP and CFP is the theory regarding managers' fiduciary duty as laid forth by Milton Friedman in his well-known article "The Social Responsibility of Business Is to Increase Its Profits" (Friedman 1970). Friedman describes that the main responsibility of managers should be to increase shareholder value, and excess investments in social projects and activities is likely to be value destructive for the firm. Henderson (2001) argues that firms investing in CSR run high risks of excess costs thus affecting profits, for example due to new required accounting systems. In an early study, Vance (1975) re-examines the abovementioned study by Moskowitz (1972) over a longer time period and finds that the overwhelming majority of the selected socially responsible firms actually performed worse than the market indices.

Lopez et al. (2007) analyze the relationship between CSP and CFP using accounting-based measures. The sample analyzed consists of two groups with 55 firms in each subsample during the years 1998-2004; one group with firms included in the Dow Jones Sustainability Index (DJSI) and the other group with firms not included in the DJSI but included in the Dow Jones Global Index (DJGI). In order to improve the degree of homogeneity of the sample, the regression model includes control variables for size, activity sector and risk. Mainly looking at the growth of profit before tax (PBT) measured by the growth in revenue, a negative relationship between CSR and accounting-based financial performance is found in the study. Lopez et al. argue that CSR related expenses can put socially responsible firms in an economic disadvantage in the short-term, and that governments through legislation and subsidies could potentially play an important role in encouraging CSR related practices among firms seeking to avoid the economic disadvantage.

Baron et al. (2011) analyze the relationship between CSR and financial performance using a sample of 1,600 firms in the US over the period 1996-2004. The KLD database is used to measure CSR score and Tobin's Q as the measure for financial performance. The study includes firm fixed effects in the estimation model in order to capture any unobserved heterogeneity in the sample and also a lagged variable approach to check the robustness of the model. The study concludes no correlation between CSR and financial performance, but a negative correlation between the social component and financial performance. The authors argue that the negative correlation may be a result of the effects on firm reputation, brand equities and productivities that social pressure may cause.

2.3 Neutral or ambiguous relationship between CSP and CFP

Some empirical research find no evidence at all of a relationship between CSR and financial performance. McWilliams and Siegel (2000) in an extensive study link data from Compustat to KLD and analyze how CSR affects financial performance over the period 1991-1996 for a total of 524 firms. The study uses a similar statistical model to Waddock and Graves (1997) but also includes a component that captures firm-investments in R&D, which has been suggested in previous literature (e.g. Griliches, 1979). The study concludes that CSR has no impact on financial performance and also emphasizes the importance of controlling for R&D in the econometric model.

Aupperle et al. (1985) examine the relationship between CSR and financial performance. In an attempt to improve the methodology of previous similar studies in the field, the authors survey CEOs using a constructed survey instrument in order to establish the corporates' social-responsibility orientation. The components in the measure of CSP in the study are economic, legal, ethical and discretionary, and a risk-adjusted ROA is used as a proxy for CFP. The authors conclude no relationship between CSP and CFP but also find that firms with a CSR committee in the corporate board do not financially outperform firms without this board-feature.

Garcia-Castro et al. (2010) address the issue of endogeneity in their study and argue that endogeneity, when not properly accounted for, provides false results and thus incorrect conclusions regarding the CSP-CFP relationship. The authors examine the CSP-CFP relationship using panel data for 658 listed firms in the US included in the KLD database since 1991. In terms of estimation methods, a combination of OLS, fixed effects and instrumental variables are used in order to enable comparison with other studies and at the same time addressing endogeneity. As a proxy for CSP, KLD data is used and as dependent variables, Tobin's Q, MVA, ROA and ROE are used. The regression model includes controlling for risk, sales, R&D and leverage. The authors conclude that when using a fixed effects model ROE, ROA and Tobin's Q support a neutral relationship between CSP and CFP. A neutral relationship is also found when the instrumental variables model is used for all of the different CFP measures. When MVA is used, a slightly negative relationship can be concluded. Lastly, the authors stress the importance of accounting for endogeneity in future studies as it might play a deciding role in terms of the nature of the CSP-CFP relationship.

2.4 Validity of different CSR ratings

Amongst previous empirical research described in **sections 2.1-2.3**, various types of CSR ratings have been used to measure CSP (e.g. KLD, ASSET4 and DJSI). As described by Ullmann (1985), the conflicting results among studies regarding how CSR affects financial performance is partly due to difficulties in measuring CSR as well as due to shortcomings amongst different CSR-measuring databases. Yet, however, there is surprisingly little research examining how well different CSR ratings converge and thus how comparable different previous empirical studies are with one another. Schwab (1980) mentions that the lack of validity in research of measures used in studies is a reason why organization studies have not made faster progress.

Chatterji et al. (2015) attempt to establish the extent to which several major CSR raters converge. Specifically, the authors analyze the convergence of KLD, FTSE4Good, DJSI, Innovest, Calvert and ASSET4 using data from 2002 to 2010. After first examining company overlaps among the six CSR raters, tetrachoric correlations between the CSR raters are performed to further assess the convergent validity. The authors use the term "convergent validity" to express the similarities between different databases. Chatterji et al. (2015) reason that two central mechanisms drive the degree of convergence between CSR raters; namely the extents of theorization and commensurability. Theorization refers to what the different CSR raters actually measure, i.e. if their theorizations of CSR are similar. As CSR is in its essence rather intangible and vaguely defined, the theorization aspect is a central issue when trying to understand and examine any potential differences between CSR raters. Commensurability refers to how different CSR raters measure the same thing. As an example the authors explain that a high degree of commensurability can occur if CSR raters measure the degree of employee safety within a firm in a similar fashion. Adjustments are made to account for differences in theorization of CSR across the six raters. Findings in the paper include little correlation between the six raters as well as little overlap in membership across the raters. Furthermore, the results suggest that the raters measure CSR components in different ways and thus that the commensurability among CSR scores is low. The authors argue that the results should signal caution when interpreting results from other studies using only one or a limited amount of raters when measuring CSR and call for further research of the validity of CSR raters.

Sharfman (1996) in his study attempts to validate the KLD measurement ratings by using a methodology known as construct validity, meaning that the KLD ratings are correlated against other types of CSR ratings. In the study, KLD is correlated against ratings generated by Fortune magazine from a corporate reputation survey as well as against a constructed list of mutual funds perceived to be socially responsible. The correlations indicate that KLD to some extent measure similar aspects of CSP as the Fortune ratings and the constructed mutual funds list and thus the authors argue that this encourages use of the KLD score.

2.5 Differences in CSR across industries and legitimacy theory

Roberts (1992) in his study examines how firm strategy impacts CSR disclosure. Moreover, in the study, measures of economic performance, strategic posture toward social responsibility as well as the power of stakeholders are used in order to assess any variations in CSR disclosure across industries. The dependent variable in the study intended to be a proxy for CSR disclosure is based on an analysis by the Council on Economic Priorities in 1986 covering over 130 firms. Roberts divides seven industries covered in the sample into two groups: high profile industries and low profile industries. The high profile group includes the automobile, airline and oil industries and are chosen due to their high levels of political risk, consumer visibility as well as intense industry competition. Roberts hypothesizes and finds that firms in the industries that belong to the high profile group are more prone to disclose CSR activities.

Campbell et al. (2003) analyze to what extent attempts of firms to construct legitimacy and thus close the legitimacy gap determines the amount of voluntary CSR disclosures. The authors of the study argue that attempts to construct legitimacy may vary depending on industry as different industries will be associated with different connotations in society at large. Firms in industries with negative associations should according to the legitimacy theory be expected to disclose more CSR related information and as an example, the authors bring up the tobacco industry which in society is by many seen as a “sinful” industry due to the harmful nature of the products. Moreover, Campbell et al. reason that for an industry as “sinful” as the tobacco industry, the goal of legitimacy strategies might not be the same for firms in industries with more positive connotations. They argue that for tobacco companies, the goal of legitimacy strategies might be to lessen negative connotations associations in society as among some stakeholders it would be impossible to completely achieve legitimacy. Using data from corporate reports and five companies that were listed in the FTSE 100 during the period January 1974 to June 1988, findings interestingly enough show that more sinful firms in the sample do not always disclose more CSR information and that in some cases the less sinful firms actually disclosed more. Possible explanations according to the authors include that some firms might not view CSR disclosure as a means of closing the legitimacy gap, and that some firms might regard CSR disclosure as an ineffective way of attaining higher legitimacy among stakeholders.

2.6 Contribution to existing literature

This paper contributes to the field of CSR in several ways. Firstly, by examining a possible link between CSP and CFP, we contribute with additional findings regarding the nature of this relationship. During the course of writing this paper and reviewing previous literature, it has become clear to us that findings relating to the CSP-CFP relationship will perhaps always be mixed as the issues suggested by Ullmann (1985) are difficult to erase entirely. Thus, any findings and conclusions from previous studies should not be looked upon in isolation – but rather together with other previous studies as results clearly vary depending on factors such as what empirical database and model is used in a study. Secondly, this paper contributes to existing literature by comparing and analyzing the KLD and ASSET4 databases in a way that, to the best of our knowledge, has not been done before. By looking at bivariate correlations as well as comparing the results from our main model using both of these two empirical databases, we are able to present evidence that support arguments that there are substantial differences between KLD and ASSET4 with regards to their theorization and commensurability. Thirdly, this study delves deep into differences relating to the benefits of CSR in terms of CFP for different industry types, based on their degree of pollution, using the framework of legitimacy theory. This analysis is important as it helps to understand the mechanics of the CSP-CFP relationship and how it differs across different industries.

3 Addressing Endogeneity

3.1 The basic problem of endogeneity

In its most basic form, the problem of endogeneity arises when an explanatory variable is correlated with the error term of a model (Wooldridge, 2013). In financial economics, the issue is well researched and is usually taken into account by using sophisticated econometric techniques to (partially) correct for endogeneity (see for example Greene, 2002). Garcia-Castro et al. (2010) note that strategic management research, particularly literature that deals with the CSP-

CFP link, has insofar been heavily skewed towards pure statistical modelling, and generally does not allude to econometrics to account for endogeneity. Endogeneity causes a serious concern for researchers, and several studies imply that it is most probably a contributing factor to the heterogeneous results in CSP-CFP literature (Jo & Harjoto, 2011).

To illustrate the issue of endogeneity, we define a simple (time-constant) OLS model which we assume will yield a true estimation of the dependent variable CFP_i:

$CFP_i = \beta CSP_i + \gamma Z_i + u_i$		**EQ.1**
CSP_i	is an observable proxy for CSP	
CFP_i	is an observable proxy for CFP	
Z_i	is a time-invariant unobservable explanatory variable (... or strategic characteristics)	
u_i	is the error term	

An assumption of **EQ.1** is that values of the explanatory variables are uncorrelated with the error term. It follows that a correlation is not directly measurable, as running the model only produces estimates of the error terms. Thus, the model can produce biased estimates. More specifically, endogeneity can commonly occur as a result of omitted-variable bias (OVB) or when there is a causal loop between the dependent and explanatory variables within a specified model (reverse causality). We now move on to discuss these two issues in the context of CSP-CFP literature.

3.2 Omitted-variable bias

Continuing the analysis above, we assume that Z_i is a confounding variable, i.e. is correlated separately with CSP_i and ε_i. Since we are actually running the regression in absence of Z_i, we will estimate:

$CFP_i = \beta CSP_i + \varepsilon_i$		**EQ.2**
ε_i	is the composite error term, i.e. $\varepsilon_i = \gamma Z_i + u_i$	

In this sense, the error term with absorb Z_i. If the results show that there is in fact a correlation between CSP_i and Z_i, and Z_i simultaneously affects CFP_i (i.e. $\gamma \neq 0$), then consequently CSP_i is correlated with ε_i, in direct violation with the aforementioned OLS assumption. It follows that the model suffers from an endogeneity problem.

In the context of strategic management, Hamilton and Nickerson (2003) argue that OVB, and as such endogeneity, arises from the fact that managers make strategic decisions based on expectations of how their decisions will affect performance of the firm in the future. This contradicts the assumption that decisions are taken randomly, which is common in most cross-sectional regression models in this field (Garcia-Castro et al., 2010). Nevertheless, if we assume that the managers build up their expectations based on internal factors (e.g. personal values and corporate culture) that are less tangible to other stakeholders, there is a basis for information asymmetry. These internal factors are either unknown or incredibly difficult to quantify to researchers. More importantly in the context of research, if any of these internal factors are separately correlated with both CFP and other strategic dimensions of a firm, then they are by definition omitted variables.

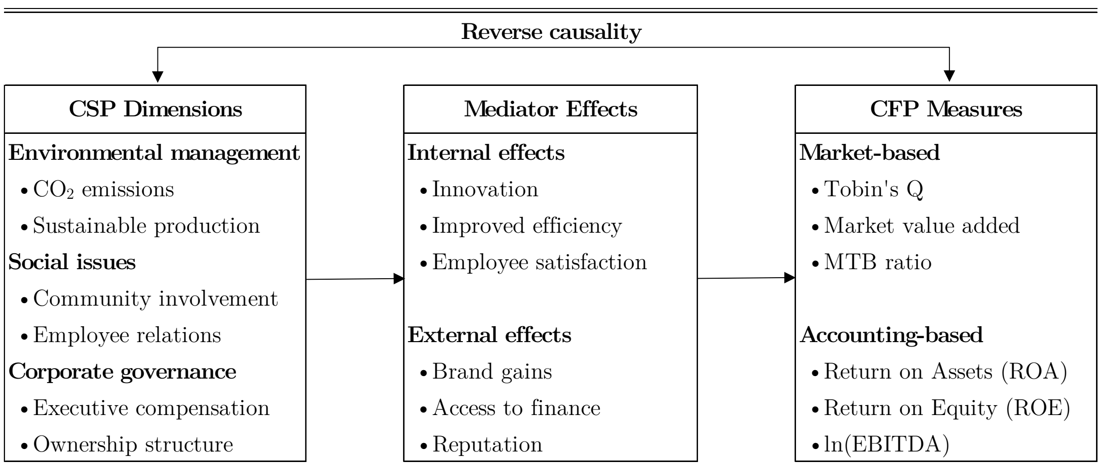
How does one account for OVB? We will now present the two prevailing methods. Firstly, several studies have tried to reproduce previous results by adding more control variables to the existing model. McWilliams and Siegel (2000) argue that when R&D intensity is added to a vast array of CSP-CFP models, the sign and significance on the CSP coefficient changes radically. The rationale behind adding R&D intensity to the model is because it has previously been shown to correlate with financial performance and CSP, and therefore controlling for it reduces the effect arising from its confounding properties. Another study found that the cultural background of a CEO could affect both financial and non-financial decisions, although found it difficult to quantify as a variable (Agle et al., 1999). The issue with this first method is self-explanatory: it is virtually impossible to identify all omitted variables, and even if we could theoretically identify a majority of them, the problem of measurability remains intact.

The second set of solutions found in literature relates to the choice of model. Some studies have tried to account for endogeneity by specifying a firm fixed effects (FE) model, with mixed results. The FE estimation proves to be a useful specification given that the inference is solely bound to the strategic behavior of the same firms in the panel data (Baltagi, 2013). One substantially more complex, and potentially flawed, method is the usage of instrumental variables (IV). We will revisit the estimator and model specification in greater detail below and in our methodology part, but first we need to address the issue of causality in CSP-CFP literature.

3.3 Reverse causality and causal claims

Suppose our research question is “Does a higher CSP lead to a higher CFP?” In order to make true causal claims about the CSP-CFP link, the explanatory variable (CSP) would need to be fully exogenous, i.e. unrelated to any observed or unobserved variables in the model. As previously discussed there is probably a vast array of unobserved (omitted) variables, leading to unobserved heterogeneity (Arellano, 2003). In context of CSP-CFP, researchers generally contextualize the relationship as some dimensions of CSP (non-randomly) existing in a firm. Internal and external mediator effects then translate this into either improved or decreased CFP. The issue of reverse causality arises from the fact that we don’t necessarily know in what direction the relationship holds. This is summarized in **Figure 1**.

Figure 1: The Reverse Causality Issue



This figure illustrates reverse causality in context of corporate social performance and corporate financial performance. Partly adopted from Shreck (2011).

In literature, two contradicting theories on the CSP-CFP link are important to mention, the good management approach and slack resources hypothesis (Waddock & Graves, 1997). We discuss these in greater detail in **section 4.1**, but more importantly the two theories are by no means collectively exhaustive or mutually exclusive. It’s not unreasonable to assume a simultaneous existence of the two, leading to a bi-directional link between CSP and CFP. We can conclude that the CSP-CFP relationship is more complex than one can intuitively imagine, and moreover, can lead to inconsistent estimates if the model(s) is designed without care.

3.4 FE estimator with CSP-CFP panel data

Building on the works of Garcia-Castro et al. (2010) and Hamilton and Nickerson (2003), we will now discuss the fixed effects estimator and how it can be conceptualized in the context of our study. Assume we gather some balanced panel data over time (t) on CSP and CFP for a number of firms (i). We can re-specify the static OLS model (**EQ.1**) as follows:

$CFP_{it} = \beta CSP_{it} + \gamma Z_i + \varepsilon_{it}$	**EQ.3**
ε_{it} is the error term	

As we can see, the panel construction now allows all variables except for Z_i to vary over time. Z_i represents the non-random strategic characteristics of firms and managers that are difficult to observe or measure for researchers. In order to account for endogeneity we now relax the OLS assumption and allow Z_i to be correlated with CSP_{it}. Under the assumption of strict exogeneity which is conditional on Z_i, it follows that:

$$E(\varepsilon_{it}|CSP_{it}, Z_i) = 0, t = 1, \dots, T \qquad \textbf{EQ.4}$$

We incorporate fixed effects by introducing firm dummy variables into the regression denoted by α_i. Another computationally different way to do this is by using a within transformation in order to demean the variables in **EQ.3**, thus eliminating Z_i

$$CFP_{it} - \overline{CFP_i} = \beta\left(\mathrm{CSP_{it}} - \overline{CSP_i}\right) + \gamma(Z_i - \overline{Z_i}) + (\varepsilon_{it} - \overline{\varepsilon_i}) \qquad \textbf{EQ.5}$$

As $Z_i = \overline{Z_i}$ we have now by construction eliminated the unobserved bias. If we run **EQ.5** using an OLS estimation, the estimates of β ought to be correct. Of course, the fixed effects model is not spotless in terms of accounting for endogeneity. Scholars in strategic management argue that the FE model is preferable to OLS, but note that the model produces biased estimates when changes in firm strategy are endogenous during a panel period (Hamilton & Nickerson, 2003). In such cases, it might make sense to look into instrumental variables as a complement to the fixed effects specification.

3.5 FE-IV approach

If the aforementioned assumption holds, namely that managers do not make strategic decisions randomly, then we have reason to suspect that CSP is not truly exogenous. One way in which researchers attempt to account for endogeneity is by using an instrumental variables (IV) approach. The fixed-effects estimator can sometimes be combined with an IV approach (see for example Varadharajan-Krishnakumar & Balestra, 1987; Wooldridge & Semykina, 2010). Wooldridge & Semykina (2010) discusses the FE-IV approach in the presence of unobserved heterogeneity as well as endogenous explanatory variables. The combination has, to the best of our knowledge, been used very scarcely in CSP-CFP literature. Battista Derchi (2015) found no evidence suggesting that the FE-IV approach would provide better estimates than a standard FE estimator when studying the link between CSR-linked executive compensation and CSP.

Theoretically, we can estimate CSP using a set of variables that (1) are related to CSP, and are (2) not correlated with the error terms in the main CFP model (orthogonality). If we assume that CSP can be measured for firm i at a given time t, then we can specify an equation system which takes the following form:

$$CSP_{it} = \beta_1 S_{it} + \alpha_{1i} + \varepsilon_{1it} \qquad \textbf{EQ.6}$$

$$CFP_{it} = \beta_2 \widehat{CSP}_{it} + \alpha_{2i} + \varepsilon_{2it} \qquad \textbf{EQ.7}$$

S_{it} in **EQ.6** is called a set of instruments. If these instruments are identified to be valid, **EQ.7** should yield unbiased estimates of β_2. Of course, a primary concern when using an IV approach is finding the appropriate instruments. In their article on diversification discount and endogeneity, Campa and Kedia (2002) argue that the equation system in an IV estimation easily becomes unidentified as most observable strategic characteristics of a firm are included in the main model.

A few studies have attempted to find appropriate instruments, with mixed results. Shreck (2011) used lagged CSP variables, although this can prove to be flawed by construction. If an unlagged CSP metric is correlated with CFP, then a lagged CSP metric (usually 1 or 2 years) is likely to also exhibit the same correlation, in direct violation with orthogonality condition (2). This is partly due to the fact that CSP ratings rarely change significantly year on year. Moreover, most studies do not provide appropriate tests regarding the validity of their choice of instruments and robustness of the overall model.

When using the IV approach, one has to be cautious of the problems that the methodology entails. Bound et al. (1995) found that instruments which explain a small portion of the variation in the endogenous regressor (CSP) can lead to biased estimates, even when they are weakly correlated to the error terms in the main model. Several studies have found that in the presence of weak instruments, IV estimations can lead to incorrect estimates (Hausman, 2001). While there are numerous suggestions on how to account for weak instruments or mismeasured variables, it is undoubtedly clear that the IV approach should be used in a complementary fashion.

4 Theories and Hypothesis

4.1 Theories

Whether or not a firm should be socially responsible and adopt CSR is a widely debated topic. Opponents argue that it does not lead to any short- or long-term financial benefits for a firm, and proponents argue the opposite. Central to these differences of opinions are two theories that need to be understood in order to fully place the CSR debate in the right context: the shareholder theory and the stakeholder theory. The shareholder theory, common in economics, evolves around that the main purpose of a firm should be to maximize shareholder value. Managers and executives in their corporate roles should not adhere to any perceived "social responsibilities" if this is outside of the scope of what lies in the interests of shareholders. Instead, any investments in social activities should be undertaken on an individual basis not using the firm's resources (Friedman, 1970). The stakeholder view is different from the original shareholder view in that it besides stockholders also recognizes the importance of other stakeholders in order for a firm to be successful and grow. Freeman (1984) describes different stakeholders such as customers, suppliers and employees and outlines ideal ways a firm can respond to their demands and interests. The importance of not isolating stakeholders and their impacts on a firm but instead focusing on interconnections is a central part of stakeholder theory, and this type of "helicopter perspective" has generated criticism among academics. Many organizational scholars argue that simultaneously focusing on different stakeholders and their different demands will lead to confusion among managers and executives which may lead to inefficiencies and conflicts within the organization (Donaldson and Preston, 1995; Jensen, 2001).

Further building on the importance of recognizing the needs of stakeholders other than stockholders as emphasized by the stakeholder theory, the legitimacy theory to some degree can be used to explain why firms adopt CSR practices. Mathews (1993) defines organizational legitimacy as the state a firm is in when the norm of social values of what is expected from a firm is in line with the social values that are a result of the firm's business activities. When the two value systems are not congruent, organizations risk facing reputational damage but it is however possible to construct legitimacy by in different ways improving the firm's image in society (Dowling & Pfeffer, 1975). Thus, building on Dowling and Pfeffer (1975), CSR practices among firms can be used as a means to reach legitimacy.

As seen in previous research, the findings regarding the relationship between CSP and CFP have been mixed. In order to understand the results of the causal links and the direction of the relationship, a theoretical framework that explains the direction of causality needs to be introduced. **Table 1** illustrates the different theories that are applicable depending on what type of direction of the causality that is found and if the links are positive, negative or neutral.

Table 1: Overview of Theories

Relationship	Positive	Negative	Neutral
CSP $\rightarrow$ CFP	Good management theory; Social impact hypothesis	Trade-off hypothesis	Neutrality theory
CSP $\leftarrow$ CFP	Slack resources hypothesis	Opportunism hypothesis	
CSP $\overset{?}{\leftrightarrow}$ CFP	Virtuous circle hypothesis		

This table provides with an overview of central theories explaining the different directions of the relationship between corporate social performance and corporate financial performance.

Starting with the direction of causality that implies that a high CSR causes improved CFP, the social impact hypothesis and the good management theory can be applied when there is a positive link. Although the theories are in many ways related to each other, they are somewhat different in terms of their definitions. According to the social impact hypothesis, which can be seen as an extension of the stakeholder theory, the recognition and meeting of demands of stakeholders that have stakes or claims in a firm will lead to higher CFP (Freeman, 1984; Preston and O'Bannon, 1997). Moreover, the positive relationships with stakeholders that often are a result of CSR practices may lead to better corporate reputation, which in turn may lead to higher CFP (Cornell & Shapiro, 1987). The good management theory, as argued by Schreck (2011), stipulates that high CSP and CFP are in fact the results of good management of a firm.

When the link is negative, implying that higher CSP is shown to be negatively correlated with CFP, the trade-off hypothesis can be used as the base explanatory framework. Building on theories by Friedman (1970), Preston and O'Bannon (1997) discuss this neoclassical-type theory saying that the costs associated with investing in CSR will cause a competitive disadvantage and negative impact on CFP compared to firms that are not as socially responsible. As Aupperle et al. (1985) mention, the competitive disadvantage is caused by a decrease in available firm resources due to the investments in social activities. Furthermore, the theory supports the findings of the study by Vance (1975) described in **section 2.2**.

When the direction is such that CFP causes a higher CSP, the slack resources hypothesis can be used as an explanatory tool. According to the slack resources theory, firms with higher financial performance are more likely to be able to afford adopting CSR practices (Waddock and Graves, 1997; Schreck, 2011). Thus, this resource based theory may also explain the findings of previous studies concluding that firm size which is assumed to be correlated with firm assets can significantly impact the amount of CSR investments a firm undertakes. When the link however is negative, the opportunism hypothesis can be applied. This theory, as discussed by Preston and O'Bannon (1997), is based on the assumption that managers often act in their own self-interests. According to the theory, managers may attempt to limit CSR related investments when financial performance is strong in order for themselves to make short-term private financial gains. Conversely, managers may undertake CSR investments when financial performance is weak in order to "counter" poor results (Preston and O'Bannon, 1997).

If a bidirectional relationship is found, this can be explained by the virtuous cycle hypothesis. As discussed by Lindgreen et al. (2016), synergetic forces cause the relationship between CSP and CFP to run in both directions. If the synergetic relationship is positive, it implies that stronger CFP leads to higher CSP, but at the same time the higher CSP causes higher CFP. A negative synergetic relationship would mean that high CSP leads to weak CFP, which in turn limits the firm's ability to make further CSR related investments – i.e. creating a virtuous cycle.

Finally, if a neutral link between CSP and CFP is found, an applicable theory is the neutrality theory. Due to a high number of intermediary and moderating variables when looking at the CSP and CFP relationship, a random link is assumed to exist (Ullmann, 1985; Waddock and Graves, 1997). Thus, according to this theory, there are no reasons to expect a relationship between CSP and CFP other than due to random effects.

4.2 Research hypothesis

Previous literature indicate that in the field of CSR, the results pertaining to any relationship between CSP and CFP are truly heterogeneous. Among these mixed findings are those who claim the relationship between CSP and CFP is neutral, i.e. non-significant. As discussed in **section 4.1**, the high number of intermediary and moderating variables suggest, at its best, only a random link between CSP and CFP (Ullmann, 1985; Waddock and Graves, 1997). Thus, the first hypothesis regarding the CSP and CFP relationship in line with neutrality theory is formulated as follows:

Hypothesis H1a: *There exists a neutral relationship between a firm's corporate social performance and its market-based and accounting-based corporate financial performance*

As described in **section 2.1**, some previous empirical research point to a positive relationship between company CSR scores and financial performance. Firm investments in CSR is seen as lowering investment risk for investors, whilst ignoring CSR in a firm's operations is considered a riskier investment (Feldman et al. 1997; Robinson et al. 2008; El Ghoul et al. 2011). Firms ignoring integrating CSR may face indirect costs in the form of, for example, reputational damage and litigation costs if products or services turn out to be unsafe or harmful (Waddock and Graves 1997; El Ghoul et al. 2011). Investing in socially responsible firms means that investors due to the lower perceived risk will demand a lower risk premium. This is in turn leads to a lower discount rate used in firm valuation by investors and thus results in higher company value (Hamilton et al. 1993; El Ghoul et al. 2011). Furthermore, investments in CSR could lead to competitive advantages that might potentially lead to financial improvements in that investments in CSR may enhance firm reputation as well as the ability to attract and retain high-quality employees (Albinger & Freeman 2000; Backhaus et al. 2002; Bauman & Skitka, 2012). Therefore, in order to establish a possible positive relationship the following hypothesis will be tested:

Hypothesis H1b: *There exists a positive relationship between a firm's corporate social performance and its market-based and accounting-based corporate financial performance*

On the contrary, there are those who argue that investing in CSR and being socially responsible does not lead to any net financial benefits for a firm due to the costs outweighing the benefits. According to shareholder theory, the fiduciary duty of managers and executives should be focused on maximizing shareholder value and any excess spending on social causes is likely value destructive and causes distraction from value enhancing projects and activities that the firm could otherwise undertake (Friedman, 1970; Friedman & Miles 2002). Investments and activities relating social responsibility should not be carried out by the corporation, but rather by the individuals (Friedman 1970). Also, according to the famous inept-custodian argument, managers and executives simply lack the necessary expertise and morals to make investment decisions regarding CSR activities. The hypothesis in order to examine a possible negative relationship is thus as follows:

Hypothesis H1c: *There exists a negative relationship between a firm's corporate social performance and its market-based and accounting-based corporate financial performance*

In previous research where the relationship between CSR scores and financial performance has been examined, different databases generating the CSR performance have been used. The majority of studies use CSP from either the KLD database, Thomson Reuters ASSET4 database or data compiled by Sustainalytics. The databases used to obtain CSP do not only vary among academic researchers, but also among companies screening and evaluating different investment alternatives based on CSP. CSR is not easily defined and the many different components constituting CSR make it difficult to quantify. Due to the intangibility and difficulty in measuring CSR, there are risks of variations in CSP across the databases. The problem with this variation is that the difference in scores may lead to incorrect and diverging conclusions regarding the relationship between CSR and financial performance as well as what investment decision should be made. Thus, this central issue should be of major interest to both academic researchers and corporations. With access to CSR data from KLD and ASSET4 the consistency and robustness of the CSR scores will be tested. The research hypothesis is as follows:

Hypothesis H2: *There is little convergence between corporate social performance as measured by the KLD and ASSET4 databases*

Using legitimacy theory voluntary engagement in CSR can be explained by firms seeking to bridge gaps between their actions, and social values and norms in the surrounding society. When employing CSR practices a firm can be said to construct legitimacy by showing that firm activities are in line with social values, e.g. by releasing CSR reports together with other company filings. As discussed in **section 2.5,** several studies have concluded that the degree and nature of CSR activities and disclosure depend on what industry a firm belongs to (e.g. Roberts, 1992; Campbell et al., 2003). Building on legitimacy theory, a plausible explanation to the variation is that some industries are more negatively looked upon by society due to factors such as harmfulness of products produced or negative environmental impact caused by manufacturing processes common in the industry. Thus, firms in industries with negative connotations may be more inclined to adopt CSR practices as it is likely that the gaps to be bridged are wider than in industries that are associated with more positive connotations. An example illustrating this is comparing firms in the tobacco industry with firms in the restaurants and bars industry. Due to factors such as harmfulness of products and unethical aspects, tobacco firms are likely to have higher incentives to construct legitimacy than firms in the restaurants and bars industry. Questions therefore arise regarding whether or not firms in industries with negative connotations have more to gain in CFP if CSR practices are adopted by firms as construction of legitimacy could be argued to have a greater affect for firms in these industries, or also cause greater financial damage if CSP is poor (Schreck, 2011). In order to examine this, industries with a high degree of pollution has been chosen as proxies for industries with bad connotations in society. As many of the gaps for firms in high-polluting industries pertain to the environmental dimension of CSR, any influence on CFP will be analyzed through this aspect. The research hypothesis can be formulated as follows:

Hypothesis H3: *The environmental dimension of corporate social responsibility will have a higher impact on corporate financial performance for firms in industries with a higher degree of pollution than in other industries.*

5 Data and Methodology

In the following section, the methodological approach and econometric model will be described in great detail. In **section 5.1** we thoroughly present the data and sample collection process. In **sections 5.2** the variables in the regression model will be laid forth and motivated. The regression specification with an exact description of the methodology will follow in **sections 5.3** and **5.4** in order to facilitate and encourage replications of this study. Lastly, we discuss our robustness checks in **sections 5.5** and **5.6**.

5.1 Data and sample collection

As this study includes a comparison of the CSP-CFP relationship between when using KLD and ASSET4 for listed companies in the US, this subsection is divided into two parts: the first part explaining the data and sample collection using KLD, and the second part explaining the process using the ASSET4 database.

When using KLD, we started by collecting CSR score data for listed firms in the US. The KLD database contains performance indicators pertaining to environmental, social and governance aspects of firms. Each year, the database is updated and several "subaspects" pertaining to the ESG dimensions may be added or removed. For reasons relating to compatibility with ASSET4, which will be explained below, and due to low variation in added or removed subaspects in the KLD database which enables comparison between years, data is collected for the period 2007-2014. Furthermore, this time period in line with previous research is deemed sufficiently long for any effects of CSR investments to "materialize" and potentially affect financial performance. Once the data was gathered, adjustments in the sample were made to remove any outliers with exceptionally high or low market capitalization and thus minimize the sample selection bias.

This is in line with previous studies stating that firm size can affect the amount of CSR related investments that is undertaken (e.g. Waddock and Graves, 1997).

When using ASSET4, ESG data for listed firms in the US was gathered. The ASSET4 database is part of Thomson Reuters and today contains data for over 5000 global companies across over 400 different ESG metrics. Similar to KLD, data is collected through publicly available information and is available starting 2002. Once again, outliers with exceptional market capitalizations were removed from the sample.

When data had been gathered for both KLD and ASSET4, adjustments were made to arrive at a final single sample. Firstly, we collect financial data from Thomson Reuters Datastream, and additional firm data from proxy statements. Secondly, we merge this with the ESG data from ASSET4 and drop all observations that do not have all firm and year observations. In line with Chatterji et al. (2015), to optimize and enable the comparability between KLD and ASSET4 we chose to drop the firms that did not overlap in the two databases. As such, we arrive at a sample of 3200 (400 U.S. listed firms) observations with all firm and year data. This takes the form of balanced panel data. **Table 2** illustrates the adjustments made in order to arrive at the final sample. Lastly, Tobin's Q, total assets, ROE, EBITDA, and debt to assets were winsorized at the 5% level in each tail in order to remove outliers. An example of an outlier was Sears Holding Corporation with an extremely negative ROE of -420% at a single point in time.

Summary statistics of the final sample is illustrated in **Table 3**. As seen by looking at the statistics in the table, the sample is skewed toward larger firms. As an example, the mean of total assets is 28,200 ($ million), the median is 12,200 ($ million) and the 75th percentile is 30,800 ($ million). With regards to net sales, the median is 4,700 ($ million) and the mean is 13,000 ($ million). The standard deviations are high for both total assets and total sales, measuring 39,600 ($ million) and 38,500 ($ million). As will be discussed later on in this paper, this is linked to the high amount of financial firms in the sample with large balance sheets.

In order to assess the industry distribution among the firms in the final sample, a supersector ticker by ICB was used. When looking at industries with greater concentration in the final sample, 14% of the firms belong in Industrial Goods and Services, 9% in Technology, 9% in Oil & Gas, 8% in Retail and 8% in Health Care. Firms included in the Financial Industry (i.e. Banks, Financial Services and Insurance) make up 19% of the sample. The industry distribution is illustrated in **Table 4**.

5.2 Variable description and summary statistics

Tobin's Q

The main focus when measuring financial performance in this study lies on the market-based measure Tobin's Q. Focusing on market-based performance indicators when assessing how CSR affects financial performance is appropriate as it means looking at the value that is assigned by the market where all forms of expectations are included (Schreck, 2009). Furthermore, the market value includes intangible values, which partly stem from a firm's investments in CSR (Manos & Drori, 2016). Fundamentally, Tobin's Q revolves around the hypothesized idea that the market value of a firm should be equal to the total value of the replacement costs, and thus a low value of Tobin's Q indicates that a firm is undervalued and a high value indicates that stocks are overvalued. Tobin's Q, a forward looking measure, captures the intangible value of a firm assigned by the market and previous literature have suggested several ways to compute the performance measure (Guenster et al., 2011). In this study, the following formula has been used to obtain Tobin's Q's for the firms in the sample:

$$Q = \frac{Market\ Capitalization + Total\ Liabilities}{Common\ Stock + Total\ Liabilities} \qquad \textbf{EQ.8}$$

In the regression model, the performance indicator is logarithmically transformed in order to normalize the distribution of residuals. Tobin's Q as a market-based performance measure has been used frequently in previous CSR related studies (e.g. Garcia-Castro et al., 2010; Schreck, 2011; Baron et al., 2011). The extensive use of Tobin's Q enables comparison between the results of this study with previous research using the same performance indicator and similar methodology.

ROE

In this study, the accounting-based performance indicator used is ROE due to the high number of previous studies using ROE as a proxy for financial performance (e.g. Waddock and Graves, 1997; Garcia-Castro et al., 2010) and due to a high number of financial companies in the sample. As the main focus lies on the market-based performance indicator Tobin's Q, ROE in this study is used to provide a more nuanced picture of the CSP to CFP relationship in order to examine whether or not the results of this study depend on what performance indicator is used in the econometric model. Thus, ROE as a performance measure also serves as a robustness check of the findings in this paper.

CSP

As previously mentioned, CSR ratings from the KLD database and Thomson Reuters ASSET4 are used in this study to measure CSP. Below, KLD and ASSET4 are described together with a summary of constructional and methodological differences between the databases. Please see **Figures 2-3** for additional information on the data construction and methodologies behind KLD and ASSET4.

KLD

KLD, or as of recently MSCI ESG KLD STATS, is a database updated annually containing positive and negative environmental, social and governance performance indicators divided into subaspects (e.g. operational waste and regulatory compliance). Each subaspect is then labelled as either a strength (if positive aspect) or concern (if negative aspect). If a firm fulfils the criteria for a strength or concern this will be marked with a "1", and if the criteria are not met this will be indicated by a "0". In order to arrive at an aggregated score for each dimension, the net of all strengths and concerns is calculated. As an example, a firm with two fulfilled strengths and one fulfilled concern in the environmental dimension will result in an aggregated environmental score of 1 (2-1). The total aggregated score across all dimensions is given by simply adding the net scores of each dimension, which in this study is how the variable CSP [KLD] is created. The final variable has a value range of -10 to 19.

The data is collected by a team of over 140 analysts and is yearly updated and assessed by looking at for example company filings and public documents, company websites and macro data from governments, NGOs and academics. Data in the KLD database is available from 1991 and today covers five different data set universes where companies are sorted depending on what index they belong to.

ASSET4

ASSET4 is a product from Thomson Reuters and contains scores ranging from 0-100 in four subdomains (Environmental, Economic, Social and Corporate Governance). Data is available from 2002 and is collected by a team of analysts looking at publicly available information relating to over 400 different ESG metrics. In order to allow for comparability with KLD, we create an equally weighted score from the ESG domains, which is the basis for our CSP [ASSET4] variable.

Control variables

In order to mitigate the potential bias due to omitted variables, several control variables are included in the regression model used in this study. Data is gathered using Thomson Reuters Datastream and complemented with proxy statement verification. Firstly, in line with previous studies (e.g. Lopez et al., 2007; Baron et al., 2011), size is controlled for using a firm's total assets which in this study is transformed logarithmically. Size has been argued to affect the financial performance of a firm (e.g. Garcia-Castro et al., 2010) and is therefore a suitable control variable.

Additionally, although not related to the main rationale behind the selection of the independent variable, it has been argued that firm size due to factors such as resource constraints and scale of operations can influence a firm's CSR activities, and previous research has shown that the motivational level to invest in CSR related projects varies depending on firm size (Waddock and Graves, 1997; Udayasankar, 2008). Even though adjustments related to size have been performed on the sample used in this study, there are still sufficient variations in total assets that motivates controlling for size. As an additional note, in contrast to studies such as Schreck (2011), sales has not been used as a proxy for size as 19% of the total sample are financial companies. The main source of revenue for banks is interest income and thus controlling for size with sales would be a misrepresentation of the sample.

Secondly, leverage as a proxy for financial risk is included as a control variable as financial leverage has been shown to affect CFP. Although debt creates tax benefits due to interest tax shields, financial distress costs may offset any tax savings depending on the level of leverage (Berk & DeMarzo, 2013). The relationship between leverage and CFP is ambiguous as previous empirical studies provide with heterogeneous results. Following previous research (e.g. McGuire et al., 1988; Waddock and Graves, 1997; Schreck, 2011), the level of debt to assets is used in this study. Furthermore, as a side note and not related to the choice of leverage as an independent variable, Dowell et al. (2000) argue that higher leverage could negatively impact a firm's ability to afford investments in CSR related projects. This is linked to the effects of debt on the cash flows of a firm.

Thirdly, cash return on assets (cash flow from operations to total assets) as a proxy for firm liquidity is controlled for in the regression model. The rationale behind including this measure lies in the fact that a high ratio indicates that cash is available for reinvestment in the firm which in turn could translate into improved financial performance. Additionally, market-to-book (lagged), Tobin's Q (lagged) and ROE (lagged) ratios are added as control variables due to their clear link to current CFP.

5.3 Regression specification

The baseline model is a two-way fixed effects model which takes the form of:

$$
\begin{aligned}
CFP_{it} &= \alpha_i + \beta_1 CSP_{i(t-1)} \\
&+ \beta_2 SIZE_{i(t-1)} \\
&+ \beta_3 RISK_{i(t-1)} \\
&+ \beta_4 LIQ_{i(t-1)} \\
&+ \beta_5 MTB_{i(t-1)} \\
&+ \beta_6 CFP_{i(t-1)} \\
&+ \lambda TIME \\
&+ \varepsilon_{it}
\end{aligned}
$$

CFP_{it}	is measured by either log(Q) or ROE for a firm *i* at time t
α_i	is a firm specific constant, i.e. the fixed effects
$CSP_{i(t-1)}$	is the aggregate score for a firms CSR measured by either KLD or ASSET4
$SIZE_{i(t-1)}$	is measured by log(Assets) for a firm *i* at time *t-1*
$RISK_{i(t-1)}$	is measured by Debt / Assets for a firm *i* at time *t-1*
$LIQ_{i(t-1)}$	is measured by Cash Return on Assets for a firm *i* at time *t-1*
$MTB_{i(t-1)}$	is measured by Market Value / Book Value for a firm *i* at time *t-1*
$CFP_{i(t-1)}$	is measured by either log(Q) or ROE for a firm *i* at time *t-1*
λ	is a set of constants to the time (2007-2014) dummies
ε_{it}	is a time-variant error term

EQ.9 specified above is a model used to examine the CSP-CFP relationship. In the model, both accounting-based and market-based indicators are used as proxies for CFP. In our analysis, we will use variations of this specification in order to test different hypotheses, as well as perform various robustness checks specified below.

We run this model first with an aggregated CSP score. After this, we perform an identical regression, but disaggregate the CSP score into the three ESG (ENV, SOC and GOV) dimensions, in order to examine if any separate domain individually has an effect on CFP. As such, we test hypotheses H1a-H1c in a collectively exhaustive manner. In line with the methodology used in previous literature, we lag our variables one year to capture the long-term nature of CSR. Practically, this means that engagement in CSR activities has been argued to take time before translating into financial impact. This is also a method to reduce distortion caused by possible reverse causality. More specifically, in the model we assume that CSP is a set of strategic decisions which then impacts CFP.

To test hypothesis H2 regarding the convergent validity of different CSP ratings, we repeat identical regression procedures for KLD and ASSET4 separately. Moreover, we compute Spearman rank correlations between the different dimensions of CSP between the two rating systems on a year-by-year basis to see if there is any directly observable convergence or divergence. The reason for using Spearman's rank correlations in place of Pearson's correlations for this study is further discussed in the results section.

Aligned with the reasoning mentioned in the **section 3** about endogeneity, we use a firm fixed effects model to account for the fact that strategic (CSR) decisions by managers are not likely to be random. Specifically, under the assumptions that unobserved variables of the model are time-invariant, we use fixed effects to describe intra-firm variation over time. We also transform the model into a two-way FE estimation by adding time dummies, in order to capture variation in CSP ratings across years. Whilst a random effects estimator might be useful for cases where there is little variation of CFP over time, the FE model minimizes the distortion caused by OVB. There is as such a case-specific explanation as to why we prefer the FE model over the RE equivalent. In order to statistically determine whether an FE or RE model should be used, we run a heteroscedasticity-robust version of the Hausman test[6]. In all our tests we can reject the null that an RE estimator is preferable to a FE estimator at a 1% significance level. Lastly, a Breusch-Pagan test is performed to assess homoscedasticity of the variables. We can reject the null of homoscedasticity in favor of heteroscedasticity at a 1% significance level (White, 1980). For the sake of coherence in this study, we use heteroscedasticity-robust standard errors across all estimations and tests.

5.4 Connecting legitimacy theory and pollution

In order to empirically test hypothesis H3 claiming that firms in high-polluting industries have more to gain in CFP from the environmental dimension of CSR, we construct dummy variable (POLLUTE) which assumes the value "1" if a firm belongs to an industry that is deemed as high-polluting or "0" if a firm belongs to an industry that is less polluting. These two groups, i.e. high-polluting industries and less pollutant industries, were constructed using the Blacksmith

[6] The alternatives to the Hausman test under non-homoskedasticity is further discussed in Chmelarova (2007)

Institute ranking as a guideline, where degree of pollution is outlined per industry. The industries in our sample that belong to the high-polluting group are auto and parts, telecom, oil and gas, chemicals, industrial goods and services, personal and household goods, health care and lastly food and beverages. In **Table 5** we see the means of the different CSP domains. A simple two-sample t-test reveals the difference is significant for the aggregated CSP score and for the SOC domain, but not for ENV and GOV.

We divide the sample according to the POLLUTE classification. In order to test the difference between the coefficients, we first follow the methodology proposed by Zellner (1962) and estimate a seemingly unrelated regression model. After storing the estimates for the four models, we run a Wald test to investigate whether the CSP coefficients are, on an individual basis, statistically equal to each other. The standard Wald statistic takes the form of:

$$W = \frac{\left[\hat{b}(CSP_{POLLUTE=1(t-1)}) - \hat{b}(CSP_{POLLUTE=0(t-1)})\right]^2}{\sigma^2\left\{\hat{b}(CSP_{POLLUTE=1(t-1)})\right\} + \sigma^2\left\{\hat{b}(CSP_{POLLUTE=0(t-1)})\right\}} \qquad \textbf{EQ.10}$$

This test statistic follows a chi-squared distribution under the null that the stored estimates of the coefficients are equal to each other, and under the assumption that the coefficients of the different groups have sampling distributions which are independent (Clogg, et al., 1995). We note that the test is sensitive to the assumption that the respective regressions of the two groups have correlated error terms. One alternative way to the test the same hypothesis would be to introduce interaction terms (e.g. ENVxPOLLUTE, SOCxPOLLUTE etc) for each of the independent variables and run the baseline regression with these. We make the judgement that the Wald test is however preferable as the 'interaction term' method may cause issues of multicollinearity.

5.5 Robustness checks

Several robustness checks have been performed in order to assess the strength and validity of the results in this study. A total of five different robustness checks are performed including an instrumental variables approach. While some robustness checks are in line with methodology used in previous studies, one of the robustness checks involves a removal of the financial firms in the sample which has, to the best of our knowledge, not been done before. The first robustness check is using ROE as an additional proxy for CFP. The second robustness check is to lag CSP in the regression model with different time periods than one year. Specifically, CSP will be lagged both two and three years. The rationale behind choosing this type of robustness check relates to the fact that it is unclear how long it takes for a firm's CSR activities to potentially translate into any impact on CFP, due to the often argued long-term and "slow" nature of CSR (Battista Derchi, 2015). Thus, it is interesting to see if any variations in the degree of lag of CSP will generate different results in terms of the CSP-CFP relationship.

The third robustness check involves using a three-year average CSR score instead of only looking at how the previous year's CSR score affects CFP. This choice of test is based on the argument that an average score may contribute to evening out any potential abnormal variation in CSR scores across consecutive years.

The fourth robustness check relates to removing the financial firms in the sample. The different nature of the financial industry compared to other industries, e.g. how revenue is earned and different CSR practices, might cause the results of the CSP-CFP relationship to be skewed toward financial firms due to their high representation in the sample.

5.6 Combined FE-IV estimation model

Building on the basic problem of endogeneity, an FE-IV estimation approach in line with that of Wooldridge & Semykina (2010) will be used as a type of robustness check, keeping in mind the limitations it poses (discussed in **section 3**).

We include all exogenous variables in the first stage in line with previous research and follow the practice to have (n+1) instruments per each (n) endogenous regressor (Larcker & Rusticus, 2010). The first instrument relates to the external pressure of a firm to adopt the practice of having a CSR report or disclose a certain degree of CSR information separate to the annual report (REPORT_PRESSURE). We measure external pressure by the proportion of peers, i.e. firms within the same industry, that have adopted this CSR practice each year. The second instrument is also related to industry-specific pressure and measures the proportion of firms that have adopted the practice of linking executive compensation to CSR metrics (COMP_PRESSURE). We can theoretically assume that industry-specific pressure correlates with a firm's CSP. Raw data on whether an individual firm has adopted the CSR reporting or CSR executive compensation practice for each given year is collected from Thomson Reuters Datastream and the variables are then constructed separately.

6 Empirical Results and Analysis

This section analyzes and evaluates the results using our models and robustness checks. It is divided on the basis of our hypotheses, and structured as follows: In **section 6.1** we present a brief overview of correlations, which is then used to draw conclusions regarding multicollinearity. In **section 6.2** we outline the CSP-CFP relationship from our main model and discuss hypotheses H1a-H1c. In **section 6.3**, we assess hypothesis H2 by comparing KLD and ASSET4 as proxies for CSP. In **section 6.4** we expand the scope of our analysis to industry-specific evidence. Specifically we cover hypothesis H3, which compares the effects of the environmental domain on CFP in high-polluting industries vis-à-vis low-polluting industries. Lastly, **section 6.5** contains the results of our robustness checks and instrumental variable analysis.

6.1 Correlation analysis

Table 6 shows correlations between unlagged (at time t) versions of the variables used in our model. We use Spearman's rank correlations *in lieu* of Pearson's due to two main reasons: Spearman's rank correlation (1) is more appropriate for *ordinal* scales, which is in line with the way KLD is measured, and (2) relaxes the assumption that the tested variables follow a bivariate normal distribution (Wayne, 1990; Chatterji et al., 2015). A variance inflation factor (VIF) analysis of our main model shows that the tolerance levels (1/VIF) are above 0.1, indicating there is probably no multicollinearity (Kutner, et al., 2004)

Firstly, while log(Q) and ROE are inherently different, it is unsurprising to find that they are significantly correlated with each other (positive, ρ_s=0.5668, $p<0.001$), as they both capture financial performance. Secondly, we find that all KLD measures (CSP, ENV, GOV and SOC) are weakly correlated (positive, $\rho_s<0.22$, $p<0.001$) with log(Q) and ROE, respectively. The same pattern is observed when using ASSET4 as a proxy for CSP. Further evaluation shows that the CSP domain which has the highest correlation with log(Q) is ENV for KLD and SOC for ASSET4, respectively. We observe an identical pattern for ROE. Additionally, as expected, the different CSP domains (i.e. SOC with GOV, GOV with ENV, etc) are positively correlated with each other, for both KLD and ASSET4. These results are expected as there are most likely CSP spillover effects for most companies. This practically means that if a company is adopting sustainable environmental practices, then the same company is likely to adopt governance and social related practices as well.

The control variables which are used in the regression show interesting patterns: SIZE is negatively correlated with both log(Q) and ROE (-0.3256 and -0.0867, $p<0.001$). This is not necessarily intuitive, as the size of a company is often argued to affect CFP positively (Lee, 2009). LIQ is positively correlated with both measures, and we find that MTB and log(Q) are highly correlated (ρ_s=0.8925, $p<0.001$), which is likely due to the fact that MTB can be seen as a proxy for the expected growth by the market. The same is observed for ROE. Furthermore, RISK is negatively correlated with ROE. This finding is interesting as one could argue that higher leverage would imply higher returns, but at the same time, leverage can have negative effects on the bottom line.

6.2 The relationship between CSP and CFP

Table 7 shows results from our main two-way fixed effects model (using KLD as a proxy for CSP) and is divided in the following way. Model (1) and (3) use the lagged aggregate CSP score as main explanatory variable for log(Q) and ROE respectively. Model (2) and (4) show the same regression, but disaggregates CSP into its three components (ENV, GOV and SOC). All control variables have been lagged one period.

While looking at models (1)-(4), we observe no significant relationship between CSP and CFP, in support of hypothesis H1a. This is true both when using an aggregated CSP score and the disaggregated components. In other words, we cannot determine whether CSP has a significant positive or negative impact on CFP. This is in line with the findings of Aupperle et al. (1985), Waddock and Graves (1997), and Garcia-Castro (2010). Possible explanations have been discussed in previous literature and theories, and will be further analyzed in the discussion. In the same models SIZE is negatively ($p<0.01$) correlated with CFP, in line with the correlation analysis. Moreover, we find that RISK is positively correlated with ROE, in line with the argument that a higher degree of leverage means that investors will require a higher return on investment to compensate for the additional risk. We fail to reach the same conclusion when using the market-based measure. No significant relationship is found in either model between LIQ and CFP. A positive relationship ($p<0.01$) is found for MTB with both CFP measures, supporting the explanation that MTB is a decent proxy for growth as implied by the market. It comes as no surprise that lagged CFP measures are positively related with unlagged CFP measures ($p<0.01$), as seen from both a market-based and account-based perspective.

From the F-tests across all models, we find a rejection of the joint null that the coefficients of our independent variables are equal to zero at 1% significance. Moreover, we find that all models using log(Q) as a proxy for CFP have a higher R^2 (explain a higher portion of variance) than the models using ROE. Similar to the arguments brought forward in Drori and Manos (2016), the reason for this could be due to the fact that market-based measures to a greater extent factor in the intangible value of CSR.

6.3 Comparing KLD and ASSET4

In order to shed light on the validity of our study and previous studies in this field, we move on to examine if we get different results when using a different proxy for CSP on the exact same sample. As mentioned before, we use ASSET4 as an alternative proxy for CSP as it is commonly used in previous research. The analysis is made up out of two parts. Firstly, we run an identical regression model using ASSET4 and present this in **Table 8**. This analysis will show if we get different results to those in **section 6.2**. Next, we look further into the differences between the CSR ratings in line with Chatterji et al. (2015), by using correlation analysis as our main tool.

In Models (1) and (3), we find no evidence of an existing relationship between CSP and CFP, when using an aggregated CSP score from ASSET4. These results are in line with what is found when performing the same analysis with KLD. However, results are different when disaggregating the CSP scores into the respective ESG domains. On the market-based side, we find a weak negative relationship ($p<0.01$) between the GOV dimension and CFP. When looking at the results using the accounting-based performance measure, we find that ENV is positively correlated with CFP. This is different to the results obtained when using KLD in that it is significant ($p<0.05$) but also positive. Moreover, the analysis using ASSET4 yields a negative relationship between SOC and CFP. In terms of the control variables, the nature of the relationships with

CFP (i.e. if positive, negative or non-existent) are the same. The implications of these findings will be discussed on the basis of theorization and commensurability in **section 7**.

By performing Spearman rank correlations between KLD and ASSET4, a clearer overview regarding the convergent validity of these two different proxies for a CSR is given. The correlations are done on a year-by-year basis in order to assess if the convergence between KLD and ASSET4 vary across years. The results from the correlation analysis is shown in **Table 9**. As seen in **Table 9**, the correlations are shown from 2007-2014 using both aggregated and disaggregated CSR scores. When looking at the aggregated version of CSP, we see that the correlations are all positive and vary across years. Between 2009 and 2010, the degree of convergence seems to have increased (i.e. from 0.2007 to 0.5798, $p<0.01$). Although it may be impossible to exactly pinpoint the underlying cause of the increase, it is possible that it is due to a higher degree of theorization and commensurability. In terms of the CSP domains, we see a rather high degree of variation across years with different significance levels. For GOV, the correlation was negative (-0.1221, $p<0.1$) in 2013.

6.4 Industry-specific evidence

We will now shift focus to industry-specific groups, in order to investigate any potential differences depending on what industry firms belong to in context of the CSP-CFP link. More specifically, building on legitimacy theory, we investigate whether firms in high polluting industries (POLLUTE=1) have more to win in terms of financial performance from being environmentally responsible, than do non-polluting firms (POLLUTE=0).

In **Table 10**, specifically in model (4), where log(Q) is the CFP measure and is regressed on the ESG components, we see that there is a negative relationship ($p<0.01$) between ENV and CFP for firms in non-polluting industries. These findings can be interpreted as companies in non-polluting industries in fact are financially worse off when investing in environmental management. The coefficients difference test results are presented in **Table 11**. We reject the null hypothesis ($p<0.05$) that the ENV coefficient in the POLLUTE=1 group is equal to the coefficient to the POLLUTE=0 group. This implies that firms in non-polluting industries have in fact more to lose when investing in environmental management than do firms in polluting industries.

6.5 Robustness checks - treatment of sample

As previously mentioned, several robustness checks have been performed in this study to reinforce the validity of the results. Firstly, we use ROE as a robustness check as another proxy for CFP. The second robustness tests involves lagging the measure of CSP with different time periods than one year. CSP was lagged both two and three years in order to examine if any differences relating to the CSP and CFP link is found. In **Table 12** the results of this robustness check are presented. We find that overall there is no major difference in the CSP to CFP relationship when lagging with two or three years[7]. We do, however, find that the sign and significance of the GOV domain shifts from non-existent to a weak and negative one ($p<0.1$). This result is in line with the results generated when using the ASSET4 scores as a proxy for CSP. This supports the argument by Battista Derchi (2015) that it is unclear how long it takes for CSP to have any financial impact. Furthermore, when lagging with two and three years, we observe that the effects of the environmental dimension on CFP becomes even greater in terms on financial benefits. In **Table 13**, we find that for the POLLUTE=1 group, ENV is now positively correlated with the market-based measure (0.0088, $p<0.05$) in contrast to the non-significant relationship seen in **Table 10**. For the POLLUTE=0 group, ENV is still negative (-0.0138, $p<0.05$). The ENV dimension remains unchanged with regards to the nature of the relationship with the accounting-based measure (ROE). A SUR-based Wald test confirms the difference in the ENV coefficient between the POLLUTE=1 and POLLUTE=0 groups.

The third robustness test involves using a three year average CSP score in order to minimize the impact on the CSP and CFP relationship caused by any abnormal variations in CSP scores across years. The results of this robustness check yields no different results than when the one-year lagged version of CSP is used. This strengthens the validity of our main results.

The fourth robustness check entails a removal of firms in the financial industry. The results are presented in **Table 14**. For the market-based measure log(Q), no differences with regards to the relationship between the CSR dimensions with CFP are found. However, with ROE we observe two differences. Firstly, the ENV dimension becomes positively correlated with CFP (0.7442, $p<0.01$). Secondly, the GOV dimension also becomes positively correlated with CFP

[7] Three years is not presented.

(0.6423, p<0.05). One of the reasons why removing financial firms causes differences when looking at ROE and its relationship with the CSP dimensions might be due to the fact that banks have volatile earnings which directly feeds through to ROE (de Haan & Poghosyan, 2012). When removing these firms, ROE is thus according to this argument normalized and the volatility in returns is dampened.

6.6 Robustness checks - FE-IV approach

We finish our robustness checks by contrasting our main model with an IV-model using aggregated CSP scores from KLD, the results of which are presented in **Table 15**. The estimation includes the same control variables, as well as firm- and year fixed effects as our main model, but assumes that that CSP is an endogenous regressor. As we can see in the second stage when using log(Q) the results on the CSP-CFP link remain insignificant and the sign of CSP is unchanged, in line with our previous results.

It makes sense however to briefly assess the validity of the instruments and overall used method. As previously mentioned, a theoretically valid instrument needs to be (1) highly correlated with the endogenous regressor (CSP), and (2) uncorrelated with the error terms in the main model (orthogonality condition). Whilst we cannot directly test these assumptions in a perfect manner, we can evaluate the outcome with common IV validity tests.

Firstly, we add REPORT_PRESSURE and COMP_PRESSURE to our main model specification and run the regression (not presented), whereas we find no significant correlation between their coefficients and CFP. Secondly, we see in the table that both instruments are significantly correlated to CSP in the first stage estimation, supporting our argument that industry-specific pressure is related to strategic decisions by means of a firms' CSP. Thirdly, a common "back-of-the-envelope" practice is to compare the first-stage F-statistic to the rule of thumb value of 10 (Staiger & Stock, 1997). After doing this, we find that the F-statistic exceeds 10 at a significance level of 1%, strengthening the choice of our instruments. Fourthly, whilst we cannot test the orthogonality conditions directly, we can perform a test of over-identifying restrictions using either Sargan or Hansen's test. We choose Hansen's test as this test statistic is consistent under conditions of heteroscedasticity (Hayashi, 2000). In an overidentified context we fail to reject the joint null hypothesis that the excluded instruments are left out correctly from our model and that the included instruments are valid (p=0.9704 for the log(Q) model). Fifthly, we compute a weak identification test using the Kleibergen-Paap Wald F-statistic. This statistic is, again, preferable to Cragg-Donald's version under conditions of heteroscedasticity-robust standard errors (Kleibergen & Paap, 2006). Comparing the outcome to the study by Stock & Yogo (2005), our results exceed the critical values at approximately 20% maximal IV size. These results neither accept nor reject the null that the excluded instruments are (weakly) correlated with the endogenous regressor (CSP). Sixthly, we compute an endogeneity test which signals a non-rejection of the null that CSP can be treated as exogenous. Collectively, whilst the IV validity test results are mixed, we reach the conclusion that there is no indication that FE-IV would provide better estimates in modelling the CSP-CFP link.

7 Conclusions and Discussion

In this study, the relationship between CSP and CFP is examined. The relationship is examined using two empirical databases; KLD and ASSET4. When using the main model with KLD, no relationship is found regardless if the market-based performance measure Tobin's Q or the accounting-based performance measure ROE is used. This supports hypothesis H1a claiming there is a non-existent relationship between CSP and CFP. This finding supports the neutrality hypothesis which stipulates that there is, at best, only a random relationship between CSP and CFP due to the many intervening and moderating variables. The finding of no relationship is in line with previous research such as Waddock and Graves (1997) and Garcia-Castro et al. (2010).

When performing the same regression using the main model but with ASSET4 as a proxy for CSP, no relationship between the aggregated CSR score and CFP is found. When disaggregating, a weak negative relationship is found between the governmental dimension and CFP measured by log(Q). When using ROE as a proxy for CFP, the environmental dimension is positively correlated with CFP and the social dimension is negatively correlated with CFP. These findings are too mixed in order to draw any conclusions, but what can be seen is that the results vary depending on if a market-based or accounting-based performance is used.

With regards to the correlations between KLD and ASSET4, variation across years is observed for the aggregated measure of CSP as well as for the disaggregated dimensions. The correlations seem to variate in a random manner, and this is likely to be due to variations in terms of the overlap of firms covered in both empirical databases as well as adding and removing score indicators in each domain. The lowest correlation between KLD and ASSET4 is observed for the governmental domain. The majority of correlations across years range from 0.4-0.6, indicating that the degree of convergent validity is moderate. The moderate degree of convergent validity between these two empirical databases can be interpreted in such a way that KLD and ASSET4 to some extent converge in terms of theorization. The implication of this finding is that previous and future research should preferably use several empirical databases when constructing proxies for CSP. Our results suggest that findings regarding the CSP and CFP relationship may vary depending on what empirical database is used. Moreover, the moderate convergent validity between KLD and ASSET4 should be taken into consideration when reviewing and comparing results of previous research.

Next, we investigated whether or not firms in high-polluting industries have more to gain in CFP from being environmentally responsible. When lagging one year and splitting our sample into the pollute and non-pollute groups, we found a negative relationship between the environmental dimension and CFP (log(Q)) for the non-polluting industries. This implies that it does not pay off to be environmentally responsible in terms of CFP if a firm is in a non-polluting industry. Using legitimacy theory, a possible explanation could be that there is simply no bridge to gap in terms of the firm activities and social values and norms and thus there are no financial benefits of being environmentally responsible. When we, however, lag the explanatory variables two years, the results are different and are in line with legitimacy theory. The two year lag yields a positive correlation between the environmental dimension and CFP measured by log(Q), and a slightly negative relationship between the environmental dimension and CFP for the non-polluting group. According to the legitimacy theory, this can be explained by the fact that firms in high-polluting industries have more to win by justifying their actions through CSR activities and thus showing that they are not in fact “that sinful”. Following the same reasoning as when lagging two years, this is in line with the legitimacy theory. Using these results, some light into the mechanics behind voluntary CSR disclosure can be shed.

8 Limitations and Suggestions for Further Research

Although this study has put great emphasis on employing a sound methodological approach where multiple robustness checks and an analysis regarding how the use of different CSR empirical databases affect the end results have been performed, several limitations need to be addressed. These limitations are important to keep in mind before drawing any general conclusions regarding the nature of the CSP-CFP relationship and any industry specific findings. Firstly, an important limitation is that CSR is vaguely defined and difficult to measure. Thus, theorization and commensurability should always be discussed in further research when a certain empirical database is used in order to critically evaluate if findings are specific to the database used. Moreover, linked to the fact that CSR is difficult to measure, it might be that empirical databases such as KLD and ASSET4 are not sophisticated enough to capture all CSR performance of firms. Some aspects of CSR might not be included in the databases and thereby the CSR scores do not entirely capture CSP.

Continuing on the theme of limitations relating to measuring CSR, weaknesses with KLD and ASSET4 need to be discussed as they affect the nature of the CSP-CFP relationship. Starting with KLD, an obvious weakness lies in the way the scores are constructed. The binary scoring system in which the strengths and concerns are given means that it is difficult to capture any gradual variation across years (Battista Derchi, 2015). In other words it is not possible to determine the extent to which a strength or concern is fulfiled and thus any variation between firms becomes more difficult to assess. As the KLD database is updated yearly, subcomponents may vary over time for companies (Servaes & Tamayo, 2013; Battista Derchi, 2015).

This issue of binary scores is not present when using ASSET4, which is an important argument to use ASSET4. Furthermore, both KLD and ASSET4 run the risk of covering firms differently depending on how long the firms have been included in the databases. This might cause biased results when comparing firms based on CSP.

An additional weakness with KLD and ASSET4, which also goes for the majority of other empirical databases measuring CSR, is that not enough focus is placed on weighing different aspects of CSR depending on what industry the firm belongs to when constructing CSR scores. It is reasonable to believe that a firm in a high-polluting industry should be more intensely scrutinized based on CSR aspects involving environmental management than a firm in a non-polluting industry. In other words, there is room for improvement in empirical databases to make sure that firms in different industries gets scored based on relevant CSR performance indicators.

In terms of the data and sample collection in this study, there are several limitations that need to be addressed. Firstly, our sample is limited to listed firms in the US and ideally a sample when investigating a matter such as CSP-CFP should contain a higher number of firms. However, it is important to remember that comparing firms with too high of a geographical distance could pose problems in terms of making comparability less meaningful. Different aspects such as the way CSR is integrated into businesses as well as different regulations could mean that comparing firms, for example, in the US with companies in Japan, where regulations to a higher degree call for CSR, becomes difficult (Krukowska, 2014). Secondly, this study only looks at data from 2007-2014. This limited time period might not be long enough to fully capture the nature of the CSP-CFP relationship. Thirdly, the summary statistics of our data shown in **Table 3** show that our sample is skewed towards larger firms. The high amount of financial firms in our sample is a reason why the mean of total assets is so large and the standard deviation is so high. This, however, is not necessarily an issue as larger firms often are those who invest the most in CSR related practices and are exposed to the most social pressure from society in terms of CSR (Ioannou et al., 2014).

The issue with endogeneity is a further limitation of this study. Although the FE estimation model as well as the IV model do to some extent account for endogeneity, it is impossible to entirely account for it. This, as is discussed throughout this paper, can cause biased results. It is clear that further research with improved ways of accounting for endogeneity is needed.

Finally, there are limitations relating to the main model used in this study. The first issue relates to the control variables used. For example R&D, which can be argued to play an important part in explaining the CSP-CFP relationship, has been left out. Further research is needed that includes controlling for R&D intensity and other related control variables. The second issue is that this study focuses solely on short-term measures of CFP (Tobin's Q, ROE). In order to capture the long-term effect of CSR, we encourage future studies to replicate a similar model design using forward-looking CFP measures (e.g. forecast P/E, ROE etc).

9 Tables

Table 2: Sample selection

Sample selection	N	Firms
Firms covered by ASSET4 with all firm and year observations	3560	445
Firms not covered by KLD + size and proxy statement adjustments	-360	-45
Final sample	3200	400

This table shows how we arrived at the final sample. First ASSET4 ESG data and financials are collected from Thomson Reuters. We drop all firms with exceptional market capitalization and all observations with incomplete firm_year data. Secondly, we match the remaining 3560 observations with MSCI ESG KLD STATS, leaving us with a final overlapping sample of 3200 observations and 400 firms.

Table 3: Summary statistics

Description	Summary statistics			Percentiles				
Variable	N	Mean	SD	p1	p25	Median	p75	p99
Total assets ($mn)	3200	28,200	39,600	2,187	5,211	12,200	30,800	157,000
Total sales ($mn)	3200	18,500	38,500	549	3,261	7,185	16,600	159,000
Total employees	3200	47,949	124,873	263	6,000	17,394	44,879	366,000
ROE	3200	14.72%	13.54%	-14.44%	7.53%	14.04%	22.27%	43.95%
Tobin's Q	3200	1.74	0.76	0.96	1.14	1.50	2.11	3.73
Operational income	3200	2,626	3,331	21	573	1,267	3,142	13,200
Debt-to-Assets ratio	3200	39.03	21.31	1.26	23.54	37.89	54.66	78.75
Market-to-Book ratio	3200	2.93	2.00	0.80	1.44	2.34	3.71	8.44
CSP [KLD]	3200	1.31	4.02	-7.00	-1.00	1.00	4.00	13.00
CSP [ASSET4]	3200	62.58	22.02	21.71	41.32	65.08	83.42	94.23

This table shows summary statistics pertaining to the sample used in this study. The number of firms (N), means, standard deviations (SD) as well as percentiles are listed for variables such as total assets and total sales.

Table 4: Industry summary statistics

Industry	Observations			CSP [KLD]		ENV [KLD]		GOV [KLD]		SOC [KLD]	
	N	Firms	Frequency	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Automobiles & Parts	40	5	1%	0.23	2.98	0.75	1.30	-0.20	0.56	-0.33	2.29
Basic Resources	80	10	3%	0.90	3.72	-0.15	1.65	-0.01	0.83	1.06	2.76
Utilities	200	25	6%	0.95	3.97	-0.25	1.58	-0.07	0.74	1.26	3.10
Telecom	48	6	2%	1.10	2.31	0.44	0.90	-0.44	0.77	1.10	1.63
Banks	144	18	5%	2.39	2.53	0.57	0.97	-0.53	0.92	2.35	2.19
Media	88	11	3%	0.78	3.26	0.45	0.88	-0.50	0.79	0.83	2.52
Oil & Gas	296	37	9%	-0.94	3.61	-0.55	1.35	-0.32	0.94	-0.07	2.71
Real Estate	184	23	6%	0.44	2.60	0.38	0.67	-0.34	0.66	0.40	1.89
Chemicals	144	18	5%	0.65	3.67	0.49	1.85	-0.05	0.66	0.21	2.56
Technology	296	37	9%	4.21	4.48	1.36	1.38	-0.42	0.86	3.27	3.32
Financial Services	176	22	6%	1.72	3.85	0.45	0.89	-0.53	0.88	1.80	3.12
Constructions & Materials	48	6	2%	-0.04	2.38	0.52	1.24	-0.25	0.67	-0.31	1.68
Industrial Goods & Services	440	55	14%	1.01	3.58	0.66	1.33	-0.37	0.82	0.72	2.57
Insurance	104	13	3%	1.41	3.09	0.57	0.84	-0.19	0.86	1.04	2.41
Personal & Household Goods	168	21	5%	1.63	5.09	0.99	1.51	-0.36	0.81	1.01	3.63
Retail	264	33	8%	0.63	3.87	0.52	0.95	-0.59	0.90	0.70	3.01
Health Care	256	32	8%	1.73	3.90	0.88	1.33	-0.53	0.75	1.38	2.68
Food & Beverage	120	15	4%	3.65	5.39	1.22	1.97	-0.13	0.95	2.56	3.67
Travel & Leisure	104	13	3%	0.97	3.97	0.80	0.99	-0.37	0.84	0.54	2.94
Total	3200	400	100%	1.31	4.02	0.53	1.40	-0.36	0.84	1.14	2.97

This table shows industry summary statistics for the sample used in this study. For all of the 19 industries, means and standard deviations for corporate social performance (CSP) scores are listed. All CSP scores are measured by the empirical database KLD. The means and standard deviations are also given in each of the three dimensions that make up the aggregated CSP. On a stand-alone basis, the Industrial Goods and Services industry has the highest frequency in the sample. Financials collectively make up 19% of the sample.

Table 5: Differences in means according to the POLLUTE classification

CSP measure	Observations		Aggregated	Disaggregated		
Group	N	Firms	CSP [KLD]	ENV [KLD]	CG [KLD]	SOC [KLD]
POLLUTE=1	1512	189	0.98	0.52	-0.33	0.79
POLLUTE=0	1688	211	1.61	0.54	-0.38	1.45
Difference	-	-	**-0.63*** **	**-0.02**	**0.05**	**-0.66*** **
Sample	3200	400	1.31	0.53	-0.36	1.14

This table shows the means for the industries in high-polluting industries (POLLUTE=1) and industries in low-polluting industries (POLLUTE=0) with corresponding t-tests for significance. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Spearman correlation analysis

Spearman correlation coefficients and significance levels between main variables

	Main variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1)	$CSP_{(t)}$ [KLD]	1															
(2)	$GOV_{(t)}$ [KLD]	0.4589***	1														
(3)	$SOC_{(t)}$ [KLD]	0.8973***	0.2461***	1													
(4)	$ENV_{(t)}$ [KLD]	0.6192***	0.1736***	0.3305***	1												
(5)	$CSP_{(t)}$ [ASSET4]	0.4671***	0.22***	0.3882***	0.396***	1											
(6)	$GOV_{(t)}$ [ASSET4]	0.2841***	0.1341***	0.2422***	0.2379***	0.7585***	1										
(7)	$SOC_{(t)}$ [ASSET4]	0.4621***	0.2064***	0.4056***	0.3576***	0.9345***	0.6193***	1									
(8)	$ENV_{(t)}$ [ASSET4]	0.4516***	0.215***	0.352***	0.4211***	0.9312***	0.6011***	0.8069***	1								
(9)	$\log(Q)_{(t)}$	0.1563***	0.0863***	0.098***	0.2105***	0.1267***	0.0803***	0.1442***	0.1141***	1							
(10)	$ROE_{(t)}$ (%)	0.1556***	0.1021***	0.1084***	0.1766***	0.2264***	0.1606***	0.2503***	0.1925***	0.5668***	1						
(11)	$COMP\_PRESSURE_{(t)}$	0.0989***	0.2488***	0.0421***	0.0133***	0.2111***	0.1895***	0.174***	0.1975***	-0.0172**	0.0123***	1					
(12)	$REPORT\_PRESSURE_{(t)}$	0.2569***	0.2954***	0.135***	0.2353***	0.2507***	0.1872***	0.2142***	0.2444***	0.1292	0.0943***	0.6771***	1				
(13)	$SIZE_{(t)}$	0.2483***	0.0549**	0.2648***	0.1058***	0.3599***	0.2286***	0.3259***	0.3723***	-0.3256***	-0.0867***	0.0317***	-0.0509***	1			
(14)	$LIQ_{(t)}$	0.0691***	0.085***	0.0261	0.1253***	0.1715***	0.1303***	0.1974***	0.1332***	0.6097***	0.7264***	0.0864***	0.1052***	-0.2615***	1		
(15)	$MTB_{(t)}$	0.1423***	0.085***	0.0857***	0.1908***	0.1282***	0.0802***	0.1486***	0.1121***	0.8925***	0.5787***	-0.0295***	0.1305**	-0.2437***	0.5056***	1	
(16)	$RISK_{(t)}$	-0.002	-0.0205	-0.0022	-0.0002	0.0609**	0.0431*	0.0273	0.0766***	-0.2375***	-0.1118***	0.058***	0.094***	0.1938***	-0.2769***	0.0451*	1

This correlation matrix illustrates the Spearman pair-wise correlation coefficients between the dependent, independent and instrumental variables used in this study. $CSP_{(t)}$ denotes the aggregated corporate social performance score at time *t*. $GOV_{(t)}$, $SOC_{(t)}$ and $ENV_{(t)}$, respectively, denote the environmental, governance and social dimensions that together make up the corporate social performance score. The [KLD]-mark and [ASSET4]-mark indicates which empirical database is used to measure the aggregated corporate social performance score and the disaggregated dimensions. $Log(Q)_{(t)}$ is the logarithm of Tobin's Q at time *t*. $ROE_{(t)}$ is the return on equity at time *t*. $COMP\_PRESSURE_{(t)}$ is an instrumental variable measuring the proportions of firms that have adopted the practice of linking executive compensation to CSR metrics. REPORT_PRESSURE is an instrumental variable relating to the external pressure to report firm-specific CSR activities in addition to the annual report. $SIZE_{(t)}$ is a control variable measured by the logarithm of a firm's total assets. $LIQ_{(t)}$ is a control variable measured by cash flow from operations to total assets. $MTB_{(t)}$ is a control variable measured by a firm's market-to-book ratio. $RISK_{(t)}$ is a control variable measured by total debt to assets as a proxy for a firm's financial risk. The correlation matrix suggests no multicollinearity. Moreover, the instrumental variables $COMP\_PRESSURE_{(t)}$ and REPORT_PRESSURE(t) are significantly correlated with corporate social performance in its aggregated and disaggregated form. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Table 7: Baseline regression (KLD)

Results of the two-way fixed effects estimation - using MSCI KLD as a proxy for CSP				
Dependent variable	$\log(Q)_{(t)}$		$ROE_{(t)}$ (%)	
Model	(1)	(2)	(3)	(4)
Explanatory variables				
$CSP_{(t-1)}$	**-0.0009**		**-0.0353**	
	(0.0013)		(0.0873)	
$ENV_{(t-1)}$		**-0.0042**		**0.0253**
		(0.0038)		(0.2688)
$GOV_{(t-1)}$		**0.0010**		**0.2784**
		(0.0043)		(0.2467)
$SOC_{(t-1)}$		**-0.0004**		**-0.1151**
		(0.0015)		(0.1030)
Controls				
$SIZE_{(t-1)}$	**-0.0698*** **	**-0.0701*** **	**-9.6336*** **	**-9.6005*** **
	(0.0184)	(0.0184)	(1.6073)	(1.6090)
$RISK_{(t-1)}$	**0.0002**	**0.0002**	**0.1052*** **	**0.1040*** **
	(0.0005)	(0.0005)	(0.0275)	(0.0273)
$LIQ_{(t-1)}$	**-0.0273**	**-0.0268**	**-8.2377**	**-8.4533**
	(0.1036)	(0.1038)	(7.1380)	(7.1804)
$MTB_{(t-1)}$	**0.0143*** **	**0.0143*** **	**2.8883*** **	**2.8902*** **
	(0.0053)	(0.0053)	(0.3288)	(0.3275)
$\log(Q)_{(t-1)}$	**0.3319*** **	**0.3315*** **		
	(0.0395)	(0.0396)		
$ROE_{(t-1)}$			**0.1758*** **	**0.1751*** **
			(0.0378)	(0.0378)
Regression details				
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
N	2,800	2,800	2,800	2,800
Unique firms	400	400	400	400
R^2	0.3865	0.3868	0.2365	0.2372
F-statistic	82.85***	71.20***	28.43***	24.61***
p-value FE vs RE	0.0000	0.0000	0.0000	0.0000

In this figure, the results of the two-way fixed effects estimation model using the KLD database to measure corporate social performance scores of firms are presented. Two measures for corporate financial performance are used: the logarithm of Tobin's Q at time *t* (model 1 and 2) and return on equity at time *t* (model 3 and 4). In model 1, the aggregated corporate social performance at time *t-1* is the dependent variable. In model 2, the aggregated form of corporate social performance is divided into its three dimensions which are then used as dependent variables: environmental ($ENV_{(t-1)}$), governance ($GOV_{(t-1)}$) and social ($SOC_{(t-1)}$). In model 3, the aggregated form in corporate social performance is used as the dependent variable. In model 4, the environmental, governance and social dimensions are used as dependent variables. Control variables are $SIZE_{(t-1)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-2, and lastly return on equity at time *t-1* in models 3-4. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Baseline regression (ASSET4)

Results of the two-way fixed effects estimation - using Thomson Reuters ASSET4 as a proxy for CSP				
Dependent variable	$\log(Q)_{(t)}$		$ROE_{(t)}$ (%)	
Model	(1)	(2)	(3)	(4)
Explanatory variables				
$CSP_{(t-1)}$	**-0.0004**		**-0.0145**	
	(0.0005)		(0.0272)	
$ENV_{(t-1)}$		**0.0005**		**0.0380****
		(0.0004)		(0.0191)
$GOV_{(t-1)}$		**-0.0013*****		**-0.0073**
		(0.0004)		(0.0238)
$SOC_{(t-1)}$		**-0.0002**		**-0.0570****
		(0.0004)		(0.0224)
Controls				
$SIZE_{(t-1)}$	**-0.0682*****	**-0.0723*****	**-9.5774*****	**-9.5769*****
	(0.0184)	(0.0185)	(1.5987)	(1.5829)
$RISK_{(t-1)}$	**0.0002**	**0.0002**	**0.1052*****	**0.1073*****
	(0.0005)	(0.0005)	(0.0274)	(0.0274)
$LIQ_{(t-1)}$	**-0.0232**	**-0.0191**	**-7.9552**	**-7.5775**
	(0.1040)	(0.1026)	(7.0820)	(7.0197)
$MTB_{(t-1)}$	**0.0141*****	**0.0141*****	**2.8872*****	**2.8810*****
	(0.0053)	(0.0052)	(0.3284)	(0.3300)
$\log(Q)_{(t-1)}$	**0.3329*****	**0.3288*****		
	(0.0393)	(0.0393)		
$ROE_{(t-1)}$			**0.1751*****	**0.1738*****
			(0.0377)	(0.0377)
Regression details				
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
N	2,800	2,800	2,800	2,800
Unique firms	400	400	400	400
R^2	0.3866	0.3904	0.2366	0.2401
F-statistic	84.62***	74.15***	28.59***	24.85***
p-value FE vs RE	0.0000	0.0000	0.0000	0.0000

In this table, the results of the two-way fixed effects estimation using Thomson Reuters ASSET4 to measure the corporate social performance of firms are presented. Two measures for corporate financial performance are used: the logarithm of Tobin's Q at time *t* (model 1 and 2) and return on equity at time *t* (model 3 and 4). In model 2, the aggregated form of corporate social performance is divided into its three dimensions which are then used as dependent variables: environmental ($ENV_{(t-1)}$), governance ($GOV_{(t-1)}$) and social ($SOC_{(t-1)}$). In model 3, the aggregated form in corporate social performance is used as the dependent variable. In model 4, the environmental, governance and social dimensions are used as dependent variables. Control variables are $SIZE_{(t-1)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-2, and lastly return on equity at time *t-1* in models 3-4. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: KLD and ASSET4 correlations

Results from year-by-year Spearman matrix between CSR domains				
Domain	CSP	ENV	GOV	SOC
Period	ρ_s	ρ_s	ρ_s	ρ_s
2007	0.2176***	0.0020	0.1352**	0.3313***
2008	0.2111***	0.0553	0.0520	0.3353***
2009	0.2007***	0.0913	0.0137	0.2870***
2010	0.5798***	0.5384***	0.3597***	0.4528***
2011	0.6200***	0.5453***	0.4009***	0.4880***
2012	0.5221***	0.4919***	-0.0050	0.4817***
2013	0.6188***	0.5444***	-0.1221*	0.5507***
2014	0.4687***	0.5916***	-0.0535	0.2705***
All years	0.4671***	0.4211***	0.1341***	0.4056***

This table illustrates the Spearman correlations between CSR domains from the two empirical databases KLD and ASSET4. Correlations are presented from 2007-2014. CSP denotes corporate social performance. ENV is the environmental dimension included in CSP. GOV is the governance dimension included in CSP. SOC is the social dimension included in CSP. At the bottom of the table, an average Spearman correlation during the years 2007-2014 is also listed. *** p<0.001, ** p<0.01, * p<0.05.

Table 10: Baseline regression divided according to the POLLUTE classification

Results from the two-way fixed effects model decomposed to high and low polluting industries - using MSCI KLD as a proxy for CSR								
Dependent variable	$\log(Q)_{(t)}$				$ROE_{(t)}$ (%)			
Group	Pollute = 1		Pollute = 0		Pollute = 1		Pollute = 0	
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explanatory variables								
$CSP_{(t-1)}$	**-0.0005**		**-0.0012**		**-0.0784**		**-0.0086**	
	(0.0017)		(0.0019)		(0.1249)		(0.1193)	
$ENV_{(t-1)}$		**0.0031**		**-0.0164*****		**-0.0337**		**0.1526**
		(0.0047)		(0.0061)		(0.3111)		(0.4683)
$GOV_{(t-1)}$		**-0.0043**		**0.0034**		**0.4057**		**0.0691**
		(0.0063)		(0.0060)		(0.3167)		(0.3657)
$SOC_{(t-1)}$		**-0.0010**		**0.0014**		**-0.2017**		**-0.0582**
		(0.0021)		(0.0022)		(0.1535)		(0.1408)
Controls								
$SIZE_{(t-1)}$	**-0.0622****	**-0.0607****	**-0.0733*****	**-0.0705*****	**-9.2206*****	**-9.1607*****	**-10.0727*****	**-10.0881*****
	(0.0248)	(0.0246)	(0.0257)	(0.0256)	(2.2018)	(2.1919)	(2.4259)	(2.4089)
$RISK_{(t-1)}$	**0.0001**	**0.0001**	**0.0000**	**0.0001**	**0.1601*****	**0.1592*****	**0.0633**	**0.0623**
	(0.0007)	(0.0007)	(0.0006)	(0.0007)	(0.0390)	(0.0387)	(0.0387)	(0.0386)
$LIQ_{(t-1)}$	**-0.1697**	**-0.1738**	**0.1188**	**0.1141**	**-16.7195***	**-16.6090**	**5.2982**	**5.1985**
	(0.1454)	(0.1460)	(0.1515)	(0.1494)	(10.0862)	(10.2107)	(10.4254)	(10.5458)
$MTB_{(t-1)}$	**0.0121***	**0.0118**	**0.0173****	**0.0166****	**3.4427*****	**3.4297*****	**2.2595*****	**2.2714*****
	(0.0072)	(0.0072)	(0.0078)	(0.0079)	(0.3949)	(0.3986)	(0.4756)	(0.4764)
$\log(Q)_{(t-1)}$	**0.4240*****	**0.4259*****	**0.2426*****	**0.2415*****				
	(0.0562)	(0.0567)	(0.0518)	(0.0517)				
$ROE_{(t-1)}$					**0.2211*****	**0.2203*****	**0.1161****	**0.1156****
					(0.0570)	(0.0572)	(0.0457)	(0.0455)
Regression details								
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
N	1,323	1,323	1,477	1,477	1,323	1,323	1,477	1,477
Unique firms	189	189	211	211	189	189	211	211
R^2	0.4430	0.4435	0.3446	0.3492	0.3081	0.3094	0.1900	0.1902
F-statistic	49.11***	43.48***	39.29***	33.47***	19.35***	17.28***	15.95***	14.39***

In this table, the results from the two-way fixed effects model is presented when the sample has been divided into a group with high-polluting industries, and another group with low-polluting industries. The KLD empirical database is used to measure corporate social performance (CSP) at time *t-1*. Two measures of corporate financial performance are used: the logarithm of Tobin's Q at time *t* (models 1-4) and return on equity (models 5-8) at time *t*. Control variables are $SIZE_{(t-1)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-4, and lastly return on equity at time *t-1* in models 5-8. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Wald tests

Results from the suest/Wald Test of difference between regression coefficients						
	Coefficients		Wald Test Results			
Models	Pollute = 1	Pollute = 0	Δ^2 between coefficients	Sum of squared	W	p-value (Prob>χ)*
Aggregated CSP, $\log(Q)_t$						
$CSP_{(t-1)}$	-0.00050	-0.00120	0.000000	0.00000	0.10000	0.75260
Disaggregated CSP, $\log(Q)_t$						
$ENV_{(t-1)}$	0.00310	-0.01640	0.00038	0.00408	10.73000	0.01000
$GOV_{(t-1)}$	-0.00430	0.00340	0.00006	0.00007	1.20000	0.27280
$SOC_{(t-1)}$	-0.00100	0.00140	0.00001	0.00001	2.34000	0.31060
Aggregated CSP, ROE_t(%)						
$CSP_{(t-1)}$	-0.07840	-0.00860	0.00487	0.00112	0.23000	0.63060
Disaggregated CSP, ROE_t(%)						
$ENV_{(t-1)}$	-0.03370	0.15260	0.03471	0.04130	1.19000	0.75510
$GOV_{(t-1)}$	0.40570	0.06910	0.11330	0.06345	0.56000	0.45500
$SOC_{(t-1)}$	-0.20170	-0.05820	0.02059	0.02100	1.02000	0.60130

This table shows the coefficients from a seemingly unrelated regression. Two measures of corporate financial performance (CFP) are used: the logarithm of Tobin's Q at time *t* and return on equity (ROE) at time *t*. For both performance measures, results are shown for the aggregated version of a firm's corporate social performance (CSP) and for the three dimensions making up CSP, at time *t-1*. The three dimensions are the environmental (ENV), the governance (GOV) and the social (SOC). All models include the control variables size, risk, liquidity, market-to-book and lagged versions of CFP. A Wald test is performed to test the null that each individual coefficient of the CSP dimensions are statistically equal to each other. The critical value for rejecting the null is p<0.05.

Table 12: 2 year lagged robustness check - Baseline regression (KLD)

Results of the two-way fixed effects estimation - using MSCI KLD as a proxy for CSP				
Dependent variable	$\log(Q)_{(t)}$		$ROE_{(t)}$ (%)	
Model	(1)	(2)	(3)	(4)
Explanatory variables				
$CSP_{(t-2)}$	**-0.0001**		**0.0345**	
	(0.0012)		(0.0737)	
$ENV_{(t-2)}$		**-0.0002**		**0.1001**
		(0.0036)		(0.2617)
$GOV_{(t-2)}$		**-0.0082***		**0.1888**
		(0.0043)		(0.2814)
$SOC_{(t-2)}$		**0.0018**		**-0.0229**
		(0.0017)		(0.1010)
Controls				
$SIZE_{(t-1)}$	**-0.0997*****	**-0.0991*****	**-8.4460*****	**-8.4518*****
	(0.0226)	(0.0226)	(1.8358)	(1.8426)
$RISK_{(t-1)}$	**0.0003**	**0.0003**	**0.0847*****	**0.0840*****
	(0.0005)	(0.0005)	(0.0294)	(0.0293)
$LIQ_{(t-1)}$	**0.0771**	**0.0817**	**-18.3008****	**-18.3098****
	(0.1066)	(0.1063)	(8.6490)	(8.6851)
$MTB_{(t-1)}$	**0.0127****	**0.0124****	**2.9716*****	**2.9768*****
	(0.0053)	(0.0053)	(0.3495)	(0.3478)
$\log(Q)_{(t-1)}$	**0.3806*****	**0.3824*****		
	(0.0406)	(0.0407)		
$ROE_{(t-1)}$			**0.2068*****	**0.2063*****
			(0.0451)	(0.0450)
Regression details				
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
N	2,400	2,400	2,400	2,400
Unique firms	400	400	400	400
R^2	0.3579	0.3596	0.2220	0.2222
F-statistic	73.71***	63.41***	21.14***	18.04***
p-value FE vs RE	0.0000	0.0000	0.0000	0.0000

In this figure, the results of a robustness check using a two-way fixed effects estimation model with the KLD database to measure corporate social performance (CSP) scores of firms are presented. Two measures for corporate financial performance are used: the logarithm of Tobin's Q at time *t* (model 1-2) and return on equity at time *t* (model 3-4). The explanatory variables are the aggregated corporate social performance at time *t-2* and the three dimensions that make up the aggregated CSP score: environmental ($ENV_{(t-2)}$), governance ($GOV_{(t-2)}$) and social ($SOC_{(t-1)}$. Control variables are $SIZE_{(t-1)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-2, and lastly return on equity at time *t-1* in models 3-4. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: 2 year robustness check - baseline regression according to the POLLUTE classification

Results from the two-way fixed effects model decomposed to high and low polluting industries - using MSCI KLD as a proxy for CSR								
Dependent variable	$\log(Q)_{(t)}$				$ROE_{(t)}$ (%)			
Group	Pollute = 1		Pollute = 0		Pollute = 1		Pollute = 0	
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explanatory variables								
$CSP_{(t-2)}$	-0.0014		0.0014		-0.0207		0.0493	
	(0.0016)		(0.0019)		(0.0913)		(0.1118)	
$ENV_{(t-2)}$		0.0088**		-0.0138**		0.0707		0.0884
		(0.0044)		(0.0058)		(0.2847)		(0.4813)
$GOV_{(t-2)}$		-0.0202***		0.0030		0.2713		0.1145
		(0.0057)		(0.0058)		(0.3759)		(0.4149)
$SOC_{(t-2)}$		-0.0004		0.0046*		-0.1337		0.0258
		(0.0023)		(0.0024)		(0.1343)		(0.1467)
Controls								
$SIZE_{(t-1)}$	-0.0676**	-0.0623*	-0.1266***	-0.1226***	-8.4490***	-8.4231***	-8.4941***	-8.5054***
	(0.0337)	(0.0336)	(0.0297)	(0.0295)	(2.4026)	(2.4073)	(2.8685)	(2.8252)
$RISK_{(t-1)}$	0.0005	0.0005	-0.0000	0.0001	0.1289***	0.1292***	0.0436	0.0430
	(0.0007)	(0.0007)	(0.0006)	(0.0006)	(0.0453)	(0.0452)	(0.0399)	(0.0395)
$LIQ_{(t-1)}$	-0.0038	0.0004	0.1808	0.1741	-29.3273***	-28.9953***	-3.0523	-3.1256
	(0.1622)	(0.1600)	(0.1411)	(0.1390)	(10.8693)	(10.8774)	(12.5063)	(12.5953)
$MTB_{(t-1)}$	0.0036	0.0029	0.0245***	0.0235***	3.3288***	3.3326***	2.5300***	2.5354***
	(0.0066)	(0.0065)	(0.0080)	(0.0079)	(0.5167)	(0.5179)	(0.4816)	(0.4796)
$\log(Q)_{(t-1)}$	0.5038***	0.5094***	0.2568***	0.2522***				
	(0.0506)	(0.0501)	(0.0563)	(0.0568)				
$ROE_{(t-1)}$					0.2882***	0.2857***	0.1133**	0.1135**
					(0.0614)	(0.0618)	(0.0565)	(0.0564)
Regression details								
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
N	1,134	1,134	1,266	1,266	1,134	1,134	1,265	1,265
Unique firms	189	189	211	211	189	189	211	211
R^2	0.4213	0.4297	0.3148	0.3202	0.2872	0.2879	0.1759	0.1760
F-statistic	49.77***	43.19***	32.70***	27.35***	19.35***	17.28***	15.95***	14.39***

In this table, the results of a robustness check using a two-way fixed effects model is presented when the sample has been divided into a group with high-polluting industries, and another group with low-polluting industries. The KLD empirical database is used to measure corporate social performance (CSP) at time *t-2*. Two measures of corporate financial performance are used: the logarithm of Tobin's Q at time *t* (models 1-4) and return on equity (models 5-8) at time *t*. The explanatory variables are the aggregated corporate social performance at time *t-2* and the three dimensions that make up the aggregated CSP score: environmental ($ENV_{(t-2)}$), governance ($GOV_{(t-2)}$) and social ($SOC_{(t-2)}$. Control variables are $SIZE_{(t-2)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-4, and lastly return on equity at time *t-1* in models 5-8. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

Table 14: No financials robustness check - Baseline regression (KLD)

Results of the two-way fixed effects estimation - using MSCI KLD as a proxy for CSP				
Dependent variable	$log(Q)_{(t)}$		$ROE_{(t)}$ (%)	
Model	(1)	(2)	(3)	(4)
Explanatory variables				
$CSP_{(t-1)}$	**-0.0008**		**-0.0243**	
	(0.0015)		(0.1001)	
$ENV_{(t-1)}$		**-0.0046**		**0.7442*****
		(0.0042)		(0.2282)
$GOV_{(t-1)}$		**0.0011**		**0.6423****
		(0.0051)		(0.2860)
$SOC_{(t-1)}$		**0.0000**		**-0.0247**
		(0.0017)		(0.1073)
Controls				
$SIZE_{(t-1)}$	**-0.0731*****	**-0.0734*****	**-9.6508*****	**-7.0893*****
	(0.0211)	(0.0212)	(1.8625)	(1.3133)
$RISK_{(t-1)}$	**-0.0003**	**-0.0003**	**0.1526*****	**0.1466*****
	(0.0005)	(0.0005)	(0.0304)	(0.0323)
$LIQ_{(t-1)}$	**-0.0735**	**-0.0721**	**-6.9038**	**-11.4566**
	(0.1101)	(0.1104)	(8.0768)	(8.3661)
$MTB_{(t-1)}$	**0.0154*****	**0.0153****	**2.9298*****	**2.9781*****
	(0.0059)	(0.0059)	(0.3511)	(0.3159)
$log(Q)_{(t-1)}$	**0.3576*****	**0.3569*****		
	(0.0427)	(0.0429)		
$ROE_{(t-1)}$			**0.1909*****	**0.1892*****
			(0.0443)	(0.0435)
Regression details				
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
N	2,267	2,267	2,267	2,267
Unique firms	324	324	324	324
R^2	0.4092	0.4096	0.2557	0.2199
F-statistic	86.26***	74.45***	26.81***	27.93***
p-value FE vs RE	0.0000	0.0000	0.0000	0.0000

This table shows the results of a robustness check using a two-way fixed effects model where the KLD empirical database has been used to measure the corporate social performance (CSP) of firms. This robustness check entails removing all firms belonging to the financial industry (banks, insurance and financial services), Two measures of corporate financial performance (CFP) are used: the logarithm of Tobin's Q at time *t* and return on equity at time *t*. Explanatory variables are the aggregated CSP score at time *t-1* as well as the three dimensions of CSP at time *t-1*: environmental (ENV), governance (GOV) and social (SOC). The explanatory variables are the aggregated corporate social performance at time *t-1* and the three dimensions that make up the aggregated CSP score: environmental ($ENV_{(t-1)}$), governance ($GOV_{(t-1)}$) and social ($SOC_{(t-1)}$. Control variables are $SIZE_{(t-1)}$ as measured by the logarithm of total assets, $RISK_{(t-1)}$ as measured by total debt to assets, $LIQ_{(t-1)}$ as measured by cash flow from operations to total assets, $MTB_{(t-1)}$ as measured by the market-to-book ratio, the logarithm of Tobin's Q at time *t-1* in models 1-2, and lastly return on equity at time *t-1* in models 3-4. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

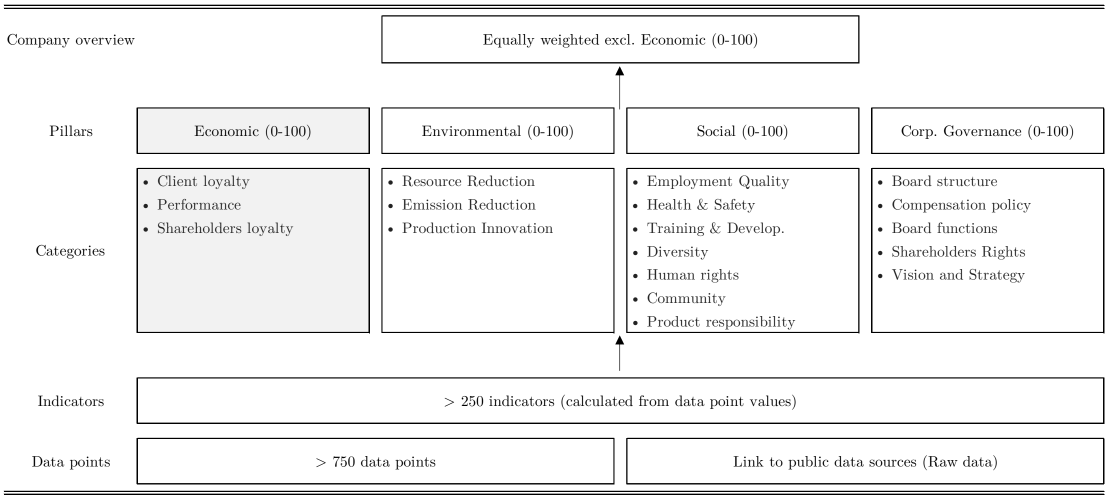
Table 15: Robustness check - FE-IV approach (KLD)

Results from Instrumental Variable regression			
Dependent variable	$\log(Q)_t$		
Estimator	FE	FE & IV	
		First stage	Second stage
Model	(1)	(2)	(3)
Explanatory variable			
$CSP_{(t-1)}$	**-0.0009**		**-0.0015**
	(0.0013)		(0.0138)
Instruments			
$COMP\_PRESSURE_{(t-1)}$		**2.9396*****	
		(0.9126)	
$REPORT\_PRESSURE_{(t-1)}$		**2.8602****	
		(1.2611)	
Controls			
$SIZE_{(t-1)}$	**-0.0698*****	**-0.6075****	**-0.0701*****
	(0.0184)	(0.2616)	(0.0198)
$RISK_{(t-1)}$	**0.0002**	**-0.0058**	**0.0002**
	(0.0005)	(0.0062)	(0.0005)
$LIQ_{(t-1)}$	**-0.0273**	**-0.5228**	**-0.0277**
	(0.1036)	(1.2524)	(0.0995)
$MTB_{(t-1)}$	**0.0143*****	**0.0537**	**0.0143*****
	(0.0053)	(0.0791)	(0.0052)
$\log(Q)_{(t-1)}$	**0.3319*****	**-0.5138**	**0.3317*****
	(0.0395)	(0.4491)	(0.0367)
Regression details			
Firm FE	YES	YES	YES
Time FE	YES	YES	YES
N	2,800	2,800	2,800
Unique firms	400	400	400
R^2	0.3865	0.3630	0.3865
F-statistic	82.85***	40.53***	106.3***
IV Validity			
Hansen Statistic			0.108
Kleibergen-Paap Wald F-statistic			9.404
Endogeneity Test Statistic			0.0112

In this table, the explanatory variable $CSP_{(t-1)}$ denotes the corporate social performance of a firm at time *t-1*. $COMP\_PRESSURE_{(t)}$ is an instrumental variable measuring the proportions of firms that have adopted the practice of linking executive compensation to CSR metrics. REPORT_PRESSURE, an instrumental variable, relates to the external pressure to report firm-specific CSR activities in addition to the annual report. Control variables are $SIZE_{(t-1)}$, $RISK_{(t-1)}$, LIQ(t-1), $MTB_{(t-1)}$ (the market-to-book ratio) and the logarithm of Tobin's Q. Model 1 is a fixed-effects estimation and models 2 and 3 represent the first and second stage of a 2SLS estimation including fixed effects. The Hansen statistic is a heteroscedasticity robust test for over identifying restrictions. The Kleibergen-Paap Wald F-statistic is compared to Stock and Yogo (2002) and tests for weak instruments. The endogeneity test is a heteroscedasticity version of the Durbin-Wu-Hausman for the null that the main regressor, i.e. CSP, can be treated as exogenous. Standard errors are robust for heteroskedascticiy. *** p<0.01, ** p<0.05, * p<0.1.

10 Appendix

Figure 2: Thomson Reuters ASSET4 methodology



This figure describes the Thomson Reuters ASSET4 methodology to arrive at a CSP score for a firm. More than 750 data points linked to (amongst others) public data sources creates more than 250 indicators, which are then used to translate into a score (ranging from 0-100) for the each CSR domains (Economic, Environmental, Social and Corporate Governance). Thomson Reuters also provides an overview score, but does not disclose how they weigh the different domains. We construct an equally-weighted CSR score by weighing Environmental, Social and Governance equally. For this study, the Economic dimension has not been included. The figure is partly adopted from Thomson Reuters.

Figure 3: MSCI ESG KLD STATS methodology

Environmental indicators	Social indicators	Governance indicators
Strengths	**Strengths**	**Stengths**
• Environmental opportunities - Clean teach	• Human Rights Policies & Initiatives	• Corruption & Instability
• Waste Management - Toxic Emissions and Waste	• Union Relations	• Financial System Risk
• Natural Resource Use - Raw Material Sourcing	• Board of Directors - Gender	
• Climate Change - Energy Efficiency	• Product Safety and Quality	
Concerns	**Concerns**	**Concerns**
• Supply Chain Management	• Support for Controversial Regimes	• Governance Structures Controversies
• Water Stress	• Labor-Management Relations	• Controversial Investments
• Impact of Products and Services	• Workforce Diversity	• Bribery & Fraud
• Operational waste	• Anticompetitive Practices	• Governance - Other Concerns

This figures describe the MSCI ESG KLD STATS methodology to arrive at a CSP score for a firm. In this figure, examples of strengths and concerns in each domain comprising corporate social responsibility are given. If a firm fulfils the criteria for a strength or concern this will be marked with a “1”, and if the criteria are not met this will be indicated by a “0”. In order to arrive at an aggregated score for each dimension, the net of all strengths and concerns is calculated. As an example, a firm with two fulfilled strengths and one fulfilled concern in the environmental dimension will result in an aggregated environmental score of 1 (2-1). The total aggregated score across all dimensions is given by simply adding the net scores of each dimension.

11 References

[1] Agle, B. R., Mitchell, R. K. & Sonnenfeld, J. A., 1999. Who Matters to CEOs? An Investigation of Stakeholder Attributes and Salience, Corporate Performance and CEO values. *Academy of Management Journal,* 42(5), pp. 507-525.

[2] Albinger, H. & Freeman, S., 2000. Corporate Social Performance and Attractiveness as an Employer to Different Job Seeking Populations. *Journal of Business Ethics,* 28(3), pp. 243-253.

[3] Arellano, M., 2003. Panel Data Econometrics. I: Oxford: Oxford University Press, pp. 7-31.

[4] Aupperle, K. E., Carroll, A. B. & Hatfield, J. D., 1985. An Empirical Examination of the Relationship between Corporate Social Responsibility and Profitability. *Academy of Management Journal,* 28(2), pp. 446-463.

[5] [Backhaus, K. B., Stone, B. A. & Heiner, K., 2002. Exploring the relationship between corporate social performance and employer attractiveness. *Business & Society,* 41(3), pp. 292-318.

[6] [Baltagi, B. H., 2013. Econometric Analysis of Panel Data (Fifth edition). I: New York: Wiley, pp. 13-63.

[7] Baron, D. P., Agus Harjoto, M. & Jo, H., 2011. The Economics and Politics of Corporate Social Performance. *Business and Politics,* 13(2), pp. 1-48.

[8] Battista Derchi, G., 2015. *Does CSR-linked Executive Compensation Promote CSR Performance?,* Lausanne: University of Lausanne.

[9] Bauman, C. W. & Skitka, L. J., 2012. Corporate Social Responsibility as a Source of Employee Satisfaction. *Research in Organizational Behavior,* Årgang 32, pp. 63-86.

[10] Berk, J. & DeMarzo, P., 2013. *Corporate Finance.* 3 red. Boston: Pearson Education Limited.

[11] Bound, J., Jaeger, D. A. & Baker, R. M., 1995. Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. *Journal of the American Statistical Association,* 90(430), pp. 443-450.

[12] Bowen, H., 1953. *Social Responsibilties of the Businessman.* New York: Harper.

[13] Campa, J. M. & Kedia, S., 2002. Explaining the Diversification Discount. *Journal of Finance,* 57(4), pp. 1731-1762.

[14] Campbell, D., Craven, B. & Shrives, P., 2003. Voluntary social reporting in three FTSE sectors a comment on perception and legitimacy. *Accounting, Auditing & Accountability Journal,* 16(4), pp. 558-581.

[15] Chatterji, A. K., Durand, R., Levine, D. I. & Touboul, S., 2015. Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers. *Strategic Management Journal,* Early View(DOI: 10.1002/smj.2407).

[16] Chmelarova, V., 2007. *The Hausman Test, And Some Alternatives, With Heteroskedastic Data,* s.l.: Graduate School of the Lousiana State University.

[17] Clogg, C. C., Petkova, E. & Haritou, A., 1995. Statistical Methods for Comparing Regression Coefficients Between Models. *The American Journal of Sociology,* 100(5), pp. 1261-1293.

[18] Cochran, P. L. & Wood, R. A., 1984. Corporate Social Responsibility and Financial Performance. *Academy of Management Journal,* 27(1), pp. 42-56.

[19] Cornell, B. & Shapiro, A. C., 1987. Corporate Stakeholders and Corporate Finance. *Financial Management,* 16(1), pp. 5-14.

[20] de Haan, J. & Poghosyan, T., 2012. Size and earnings volatility of US bank holding companies. *Journal of Banking and Finance,* 36(11), pp. 3008-3016.

[21] Diebold, F. X., 2000. *Elements of Forecasting.* 2 red. s.l.:Cengage Learning.

[22] Donaldson, T. & Preston, L. E., 1995. The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *The Academy of Management Review,* 20(1), pp. 65-91.

[23] Dowell, G., Hart, S. & Yeung, B., 2000. Do Corporate Global Environmental Standards Create or Destroy Market Value?. *Management Science,* 46(8), pp. 1059-1074.

[24] Dowling, J. & Pfeffer, J., 1975. Organizational Legitimacy: Social Values and Organizational Behavior. *The Pacific Sociological Review,* 18(1), pp. 122-136.

[25] El Ghoul, S., Guedhami, O., Kwok, C. C. & Mishra, D. R., 2011. Does corporate social responsibility affect the cost of capital?. *Journal of Banking & Finance,* 35(9), pp. 2388-2406.

[26] Feldman, S. J., Soyka, P. A. & Ameer, P. G., 1997. Does Improving a Firm's Environmental Management System and Environmental Performance Result in a Higher Stock Price?. *The Journal of Investing,* 6(4), pp. 87-97.

[27] Freeman, E. R., 1984. *Strategic Management: A Stakeholder Approach.* s.l.:Pitman.

[28] Friedman, A. L. & Miles, S., 2002. Developing Stakeholder Theory. *Journal of Management Studies,* 39(1), pp. 1-21.

[29] Friedman, M., 1970. The Social Responsibility of Business is to Increase its Profits. *The New York Times Magazine*, 13 September.

[30] Garcia-Castro, R., Ariño, M. A. & Canela, M. A., 2010. Does Social Performance Really Lead to Financial Performance? Accounting for Endogeneity. *Journal of Business Ethics,* 92(1), pp. 107-126.

[31] Granger, C. W. J., 1969. Investigating Causal Relations By Econometric Models and Cross-spectral Methods. *Econometrica,* 37(3), pp. 424-438.

[32] Greene, W. H., 2002. Econometric Analysis (Fifth edition). I: New Jersey: Pearson Education, pp. 339-425.

[33] Griliches, Z., 1979. Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics,* 10(1), pp. 92-116.

[34] Guenster, N., Bauer, R., Derwall, J. & Koedijk, K., 2011. The Economic Value of Corporate Eco-Efficiency. *European Financial Management,* 17(4), pp. 679-704.

[35] Hamilton, B. H. & Nickerson, J. A., 2003. Correcting for Endogeneity in Strategic Management. *Strategic Organization,* 1(1), pp. 51-78.

[36] Hamilton, S., Jo, H. & Statman, M., 1993. Doing well while doing good? The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysis Journal,* 49(1), pp. 62-97.

[37] Hausman, J., 2001. Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives,* 15(4), pp. 57-67.

[38] Hayashi, F., 2000. Econometrics. I: Princeton: Princeton University Press, pp. 206-213, 226-228, 407-417.

[39] Henderson, D., 2001. *Misguided Virtue False Notions of Corporate Social Responsibility.* London: The Institute of Economic Affairs.

[40] Ioannou, I., Cheng, B. & Serafeim, G., 2014. Corporate social responsibility and access to finance. *Strategic Management Journal,* 35(1), pp. 1-23.

[41] Jensen, M. C., 2001. Value Maximisation, Stakeholder Theory, and the Corporate Objective Function. *European Financial Management,* 7(3), pp. 297-317.

[42] Jo, H. & Harjoto, M. A., 2011. Corporate Governance and Firm Value: The Impact of Corporate Social Responsibility. *Journal of Business Ethics,* 103(3), pp. 351-383.

[43] Kleibergen, F. & Paap, R., 2006. Generalized Reduced Rank Tests using the Singular. *Journal of Econometrics,* Årgang 127, pp. 97-126.

[44] Krukowska, M., 2014. Determinants Of Corporate Social Responsibility In Japanese Companies. *Management and Business Administration, Central Europe,* 22(4), pp. 39-57.

[45] Kutner, M. H., Nachtsheim, C. J. & Neter, J., 2004. *Applied Linear Regression Models (Fourth edition).* Irwin: McGraw-Hill.

[46] Larcker, D. F. & Rusticus, T. O., 2010. On the use of Instrumental Variables in Accounting Research. *Journal of Accounting and Economics,* 49(3), pp. 186-205.

[47] Lee, J., 2009. Does Size Matter in Firm Performance? Evidence from US Public Firms. *International Journal of the Economics of Business,* 16(2), pp. 189-203.

[48] Lindgreen, A., Kotler, P., Vanhamme, J. & Maon, F., 2016. *A Stakeholder Approach to Corporate Social Responsibility: Pressures, Conflicts, and Reconciliation.* New York: Routledge.

[49] Lopez, M., Garcia, A. & Rodriguez, L., 2007. Sustainable Development and Corporate Performance: A Study Based on the Dow Jones Sustainability Index. *Journal of Business Ethics,* 75(3), pp. 285-300.

[50] Manos, R. & Drori, I., 2016. *Corporate Responsibility.* 1 red. s.l.:Palgrave Macmillan .

[51] McGuire, J. B., Sundgren, A. & Schneeweis, T., 1988. Corporate Social Responsbility and Firm Financial Performance. *Academy of Management Journal,* 31(4), pp. 854-872.

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

[53] Moskowitz, M., 1972. Choosing Socially Responsible Stocks. *Business and Society Review,* Issue 1, pp. 71-75.

[54] Orlitzky, M., Schmidt, F. L. & Rynes, S. L., 2003. Corporate Social and Financial Performance: A Meta-analysis. *Organization Studies,* 24(3), pp. 403-441.

[55] Preston, L. E. & O'Bannon, D. P., 1997. The Corporate Social-Financial Performance Relationship. *Business and Society,* 36(4), pp. 419-429.

[56] Roberts, R. W., 1992. Determinants of Corporate Social Responsibility Disclosure: An Application of Stakeholder Theory. *Accounting, Organizations and Society,* 17(6), pp. 595-612.

[57] Robinson, M. J., Kleffner, A. & Bertels, S., 2008. *The Value of a Reputation For Corporate Social Responsibility: Empirical Evidence,* s.l.: Working paper, University of Michigan.

[58] Schreck, P., 2009. *The Business Case for Corporate Social Responsibility.* 1 red. Heidelberg: Physica-Verlag.

[59] Schreck, P., 2011. Reviewing the Business Case for Corporate Social Responsibility: New Evidence and Analysis. *Journal of Business Ethics,* 2(103), pp. 167-188.

[60] Schwab, D. P., 1980. Construct validity in organization behaviour. *Research in Organizational Behaviour,* Issue 2, pp. 3-43.

[61] Servaes, H. & Tamayo, A., 2013. The Impact of Corporate Social Responsibility on Firm Value: The Role of Customer Awareness. *Management Science,* 59(5), pp. 1045-1061.

[62] Sharfman, M., 1996. The construct validity of the Kinder, Lydenberg & Domini social performance ratings data. *Journal of Business Ethics,* 15(3), pp. 287-296.

[63] Staiger, D. & Stock, J. H., 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica,* 65(3), pp. 557-586.

[64] Stock, J. & Yogo, M., 2005. Testing for Weak Instruments in Linear IV Regression. I: *Andrews DWK Identification and Inference for Econometric Models.* New York: Cambridge University Press, pp. 80-108.

[65] TIME, 2012. *Why Companies Can No Longer Afford to Ignore Their Social Responsibilities,* Pennsylvania: Knowledge@Wharton.

[66] Udayasankar, K., 2008. Corporate social responsibility and firm size. *Journal of Business Ethics,* 83(2), pp. 167-175.

[67] Ullmann, A. A., 1985. Data in Search of a Theory: A Critical Examination of the Relationships Among Social Performance, Social Disclosure, and Economic Performance of U.S. Firms. *Academy of Management Review,* 10(3), pp. 540-557.

[68] Waddock, S. A. & Graves, S. B., 1997. The Corporate Social Performance-Financial Performance Link. *Strategic Management Journal,* 18(4), pp. 303-319.

[69] Vance, S. C., 1975. Are Socially Responsible Corporations Good Investment Risks?. *Management Review,* 64(8), pp. 18-24.

[70] Varadharajan-Krishnakumar, J. & Balestra, P., 1987. Full Information Estimations of a System of Simultaneous Equations with Error Component Structure. *Econometric Theory,* 3(2), pp. 223-246.

[71] Wayne, D. W., 1990. Spearman rank correlation coefficient (Second edition). I: *Applied Nonparametric Statistics.* Boston: PWS-Kent, pp. 358-365.

[72] White, H., 1980. A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity. *Econometrica,* 48(4), pp. 817-818.

[73] Wooldridge, J. M., 2013. Introductory Econometrics: A Modern Approach (Fifth Edition). I: Australia: South-Western College Publications, pp. 82-83, 490-528.

[74] Wooldridge, J. M. & Semykina, A., 2010. Estimating Panel Data Models in the Presence of Endogeneity and Selection. *Journal of Econometrics,* 157(2), pp. 375-380.

[75] Zellner, A., 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggegation Bias. *Journal of the American Statistical Association,* 57(298), pp. 348-368.

12 MSCI data disclaimer

The following disclaimer pertains only to ESG data derived from the MSCI ESG KLD STATS database. In the thesis we refer to this data interchangeably as KLD or MSCI KLD:

The ESG data contained herein is the property of MSCI ESG Research Inc. (ESG). ESG, its affiliates and information providers make no warranties with respect to any such data. The ESG data contained herein is used under license and may not be further used, distributed or disseminated without the express written consent of ESG.