Revisiting the Relationship between Corporate Social and Financial Performance

Universität Hamburg
Fakultät Wirtschafts- und Sozialwissenschaften

Kumulative Dissertation

Zur Erlangung der Würde eines Doktors der
Wirtschafts- und Sozialwissenschaften
„Dr. rer. pol.“
(gemäß der PromO vom 18. Januar 2017)

vorgelegt von
Maximilian Schnippering
aus Kühlungsborn

Hamburg, den 16. Oktober 2020

Vorsitz: Prof. Dr. Frank Schiemann

Erstgutachter: Prof. Dr. Timo Busch

Zweitgutachter: Prof. Dr. Alexander Bassen

Datum der Disputation: 22.03.2021

Erklärung

Hiermit erkläre ich, Maximilian Schnippering, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Ort/Datum	Unterschrift Doktorand/in

Eidesstattliche Versicherung

Ich, Maximilian Schnippering, versichere an Eides statt, dass ich die Dissertation mit dem Titel:

„**Revisiting the Relationship between Corporate Social and Financial Performance**“

selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach § 6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

____________________________ Ort/Datum

____ Unterschrift Doktorand/in

Selbstdeklaration bei kumulativen Promotionen

Konzeption / Planung: Formulierung des grundlegenden wissenschaftlichen Problems, basierend auf bisher unbeantworteten theoretischen Fragestellungen inklusive der Zusammenfassung der generellen Fragen, die anhand von Analysen oder Experimenten/Untersuchungen beantwortbar sind. Planung der Experimente/ Analysen und Formulierung der methodischen Vorgehensweise, inklusive Wahl der Methode und unabhängige methodologische Entwicklung.

Durchführung: Grad der Einbindung in die konkreten Untersuchungen bzw. Analysen.

Manuskripterstellung: Präsentation, Interpretation und Diskussion der erzielten Ergebnisse in Form eines wissenschaftlichen Artikels.

Die Einschätzung des geleisteten Anteils erfolgt mittels Punkteinschätzung von 1 – 100 %.

Für mindestens einen der vorliegenden Artikel liegt die Eigenleistung bei 100 %.

Für einen zweiten Artikel liegt die Eigenleistung für

das Konzept / die Planung bei	50%
die Durchführung bei	50%
der Manuskripterstellung bei	50%

Für einen dritten Artikel liegt die Eigenleistung für

das Konzept / die Planung bei	25%
die Durchführung bei	50%
der Manuskripterstellung bei	50%

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit den am Artikel beteiligten Koautoren einvernehmlich abgestimmt.

____________________ Ort/Datum

____________________ Unterschrift Doktorand/in

Acknowledgements

Finalizing a PhD during a pandemic is very challenging. All over the world, people are suffering from the current situation and its consequences. I am very thankful for being in a comparable safe environment, which still offers so many opportunities in this unprecedented situation. Over the last four years, I have been given the opportunity to participate in a very special setup. I have conducted this PhD thesis while being part-time employed by the University of Hamburg and part-time by Siemens Gamesa Renewable Energy.

I would like to express my gratitude to my supervisor Prof. Dr. Timo Busch. Timo, I am very grateful to have learnt so many valuable lessons from you. You provided valuable feedback whenever you identified development potentials and you have always been available when I needed your advice. Furthermore, I would like to thank my second supervisor, Prof. Dr. Alexander Bassen, for his support and valuable feedback in many colloquia. Additionally, I would also like to express my thanks to the whole team of the Management & Sustainability chair. Especially, I would like to thank Matthew for having always an open ear and many developmental thoughts – you have not only been a great colleague, but also a great friend. The same also accounts for Rita, Sarah, Caro, Wiebke, Simone and the 1st generation PhDs. All of you have really impacted my path for not only having this PhD finished, but also fostering my personal development.

Additionally, I would also like to express my thanks to my colleagues and mentors at SGRE. Business realities sometimes severely limit the time for research, but you have always given me the freedom and encouraged me to pursue my project.

I would also like to extend my deepest gratitude to my family and pay tribute to the loved ones we have lost along the way. Specifically, I would like to thank my wife, Anja, our kids, Avienne and Gustav, my parents, Anne and Michael, as well as Inge, Axel. You have encouraged me at every step, and I am really glad to have you by my side. Finally, I would also like to thank my close friends. Following your individual paths, you have always been a great inspiration to me.

Abstract

Despite tremendous efforts in the strategic management literature, researchers are still reporting negative, neutral as well as positive results, when analyzing the relationship between corporate social performance (CSP) and corporate financial performance (CFP). Reviewing potential causes for this ambiguity, different econometrical issues have been identified, such as omitted variables, reverse causality or measurement errors. This cumulative dissertation project is focused on a critical assessment of measurement errors, examining how relevant different measurements of CSP are for the analysis of the relation between CSP and CFP.

While it has been established that CSP is a multi-dimensional construct, no standard measurement approach has emerged. Instead, various measurements of CSP have been developed differing by data source and level of aggregation. However, different measurements of CSP also lead to different results and implications. Analyzing the relevance of different CSP measurements, this project is aiming to increase the transparency on three aspects: first, the interrelations between different CSP measurements; second, the materiality of CSP improvements; and third, interferences with control variables, at the example of R&D intensity.

By means of a systematic literature review and econometrical panel analyses, three research papers have been developed. The results emphasize that different measurements of CSP significantly impact the analysis of the relation between CSP-CFP. First, improvements in disaggregated operational CSP (e.g. CO_2 emissions) are not reflected in aggregated CSP provided by rating agencies (e.g. KLD scores). Thus, disaggregated and aggregated CSP measurements are not per se related with each other and appear to cover different aspects of CSP. Second, improvements in disaggregated operational CSP are positively related with CFP. This highlights the dynamic capabilities associated with CSP and encourages current leaders as well as laggards to continuously improve CSP on the operational level. Finally, different measurements of CSP do not merely influence the relation between CSP-CFP, but also between CFP and control variables. Highly aggregated CSP measurements tend to cover similar aspects like other constructs, such as R&D intensity, while disaggregated CSP reveals that certain aspects of CSP contain financially material information and need to be actively addressed.

Accordingly, the results contribute to the literature by emphasizing the implications of different CSP measurements. Aggregated and disaggregated measurements of CSP need to be assessed carefully as they are influencing the relationship between CSP-CFP and the control variables. Different measurement approaches contain financially material information and cannot replace but complement each other.

Zusammenfassung

Obgleich großer Anstrengungen in der Managementforschung werden bis heute negative, neutrale und positive Beziehungen bei der Analyse der Beziehung zwischen Corporate Social Performance (CSP) und Corporate Financial Performance (CFP) berichtet. Ökonometrische Verzerrungen, wie ausgelassene Variablen, simultane Kausalität oder Messfehler wurden als mögliche Ursachen der uneindeutigen Ergebnisse identifiziert. Der Fokus dieser kumulativen Dissertation liegt auf einer kritischen Bewertung von Messfehlern, um zu untersuchen, inwieweit unterschiedliche Messgrößen für CSP von Relevanz sind bei der Analyse der Beziehung zwischen CSP und CFP.

Obwohl festgestellt wurde, dass es sich bei CSP um ein mehrdimensionales Konstrukt handelt, hat sich bislang kein Standardmessansatz herausgebildet. Stattdessen wurden verschiedene Messgrößen entwickelt, die sich hinsichtlich der Datenquelle und Aggregationsebene unterscheiden. Unterschiedliche Messgrößen führen allerdings auch zu unterschiedlichen Ergebnissen und Implikationen. Durch die Analyse der Relevanz verschiedener CSP-Messgrößen zielt diese Arbeit darauf ab, die Transparenz in Bezug auf drei Aspekte zu erhöhen: erstens hinsichtlich der Wechselbeziehungen zwischen verschiedenen CSP-Messgrößen, zweitens hinsichtlich der finanziellen Wesentlichkeit von Verbesserungen der CSP und drittens hinsichtlich Interferenzen mit Kontrollvariablen am Beispiel der F&E-Intensität.

Mit Hilfe einer systematische Literaturanalyse und ökonometrischen Panelanalysen wurden drei Forschungsarbeiten entwickelt. Die Ergebnisse zeigen, dass verschiedene Messgrößen von CSP die Analyse der Beziehung zwischen CSP-CFP signifikant beeinflussen. Erstens sind Verbesserungen von disaggregierten, operativen CSP-Messgrößen (z.B. CO2-Emissionen) nicht in aggregierten CSP-Messgrößen reflektiert, die von Rating-Agenturen bereitgestellt werden (z.B. KLD-Scores). Daher sind disaggregierte und aggregierte CSP-Messgrößen nicht per se miteinander verbunden und scheinen verschiedene Aspekte abzudecken. Zweitens stehen Verbesserungen von disaggregierten, operativen CSP-Messgrößen in einem positiven Zusammenhang mit der CFP. Dies verdeutlicht die dynamischen Eigenschaften, die mit CSP verbunden sind und ermutigt sowohl die Unternehmen mit der höchsten CSP als auch die Nachzügler, CSP auf operativer Ebene kontinuierlich zu verbessern. Außerdem beeinflussen unterschiedliche Messgrößen von CSP nicht nur die Beziehung zwischen CSP-CFP, sondern auch zwischen CFP und Kontrollvariablen. Hochgradig aggregierte CSP-Messgrößen tendieren dazu, ähnliche Aspekte wie andere Konstrukte abzudecken, zum Beispiel die F&E-Intensität,

während disaggregierte CSP-Messgrößen zeigen, dass bestimmte Aspekte der CSP finanziell wesentliche Informationen enthalten und aktiv adressiert werden müssen.

Die Ergebnisse tragen zur Literatur bei, indem sie die Implikationen der verschiedenen CSP-Messgrößen hervorheben. Aggregierte und disaggregierte CSP-Messgrößen müssen sorgfältig bewertet werden, da sie die Beziehung zwischen CSP-CFP und den Kontrollvariablen beeinflussen. Außerdem enthalten unterschiedliche Messansätze finanziell wesentliche Informationen und können einander nicht ersetzen, aber ergänzen.

Table of Contents

1. Introduction

Research is showing that we are at an extremely critical point in time, increasingly close to environmental tipping points that are likely to cause severe damages to our planetary systems as well as increasingly close to social tipping points that bind together the needs of different generations (Rockstroem et al. 2009; Steffen et al. 2015). Companies have been identified as a central catalyst for fostering corporate sustainability, as firms *"must apply these principles to their products, policies and practices"* (Bansal 2005, p. 199). Accordingly, research on corporate sustainability is critical in order to better understand the underlying mechanisms and to derive focused measures to mitigate environmental and social degradation.

Our understanding of corporate sustainability is rooted in many advancements made in the second half of the 20th century. Initial studies were focused on developing a common understanding of corporate sustainability as well as examining its financial materiality (Bansal 2005; Carroll 1979; Griffin 2000; Griffin and Mahon 1997; Margolis et al. 2007; Margolis and Walsh 2003; Ullman 1985). Reviewing the strategic management literature on corporate sustainability, many studies are analyzing the relationship between corporate sustainability, often expressed by corporate social performance (CSP)[1], and corporate financial performance (CFP). So-called CSP-CFP studies examine whether there is a business case for sustainability or whether sustainability is only a costly philanthropical activity (Friedman 1970). The business case for sustainability is mainly based on instrumental stakeholder theory, hypothesizing that companies with a good corporate social performance will also be financially rewarded (e.g. Allouche 2005; Brammer and Millington 2008; Donaldson and Preston 1995; Jones 1995; Orlitzky et al. 2003).

However, analyzing the relationship between CSP and CFP is not a straightforward process, which is still discussed after 40 years of research. Researchers have found negative, neutral as well as positive relationships between CSP and CFP (Friede et al. 2015; Orlitzky et al. 2003). Thus, despite tremendous efforts further research is still needed to analyze the relationship between CSP and CFP in order to better understand the underlying mechanisms and influencing factors that cause the ambiguous results. Many studies have pointed out severe econometrical issues, such as omitted variable bias (e.g. Boulouta 2013; McWilliams and Siegel 2000), functional misspecification (e.g. Barnett and Salomon 2012), simultaneity bias (e.g. Jo and Harjoto 2011) and measurement errors (e.g. Orlitzky and Benjamin 2001), as aspects that are

[1] Here, CSP will be used as an umbrella term that incorporates social and environmental aspects, unless stated otherwise.

significantly influencing the underlying results. For this cumulative dissertation project, the main research focus is directed towards a critical assessment of different CSP measurements.

From a measurement perspective, CSP can be considered a multi-dimensional latent construct. In econometrics, latent constructs are not directly observable. Accordingly, every measurement is a function of an assumed true value and an error term (Chatterji et al. 2007; Orlitzky et al. 2003). This is becoming even more complex as CSP is covering multiple ecological and social aspects (e.g. Brammer and Millington 2008; Margolis et al. 2007; Wood 2010). Various measurement approaches have been proposed, differing by data source and level of aggregation.

When considering the different sources of CSP measurements, two main sources of data can be distinguished. On the one side, companies publish CSP data by themselves and make this information externally available (e.g., via non-financial reporting). On the other side, a large industry of external rating agencies has developed, which is also providing CSP ratings based on assessments of publicly available information as well as based on company surveys (e.g., Asset4, MSCI KLD, RobecoSAM, Sustainalytics, TruCost). When looking at the different levels of aggregation, aggregated and disaggregated measurements of CSP have emerged. For aggregated CSP measurements, multiple CSP aspects are blended into a single variable to capture the multi-dimensionality and utilize an overall assessment per company (e.g., index affiliations, CSP scores). Next to that, disaggregated measurements of CSP are focusing on sub-themes to clearly allow for assessments of individual CSP pillars (e.g. environmental pillar scores) or individual data points (e.g. CO_2 emissions). However, due to the ambiguous results, the efficiency of CSP measurements has been highly criticized (Mattingly and Berman 2006). Efficient measurements need to be valid and reliable in order to mitigate corresponding endogeneity issues. Validity describes the degree of accuracy of the measurement, while reliability is referring to the "exact replicability of the […] results" (Leung 2015, p. 326).

On the one side, different measurements of CSP lead to different implications for control variables and the relationship between CSP and CFP. On the other side, being a multi-dimensional construct, different measurements of CSP are necessary to broaden the view and assess the impacts from different perspectives and levels. The aim of this project is to bridge these different measurement approaches, including data directly taken from companies' non-financial reporting and third-party aggregated assessments of CSP, and improve the transparency on the effects resulting from varying measurements in order to derive meaningful implications for managers, investors and decision makers. This project is addressing these gaps in two regards: First, by analyzing how different measurements of CSP can lead to interferences with other constructs, at the example of innovation[2] and, second, by analyzing the materiality of CSP improvements on a disaggregated level and its interrelations with aggregated CSP measurements.

Consequentially, the results will contribute to the literature by fostering the understanding on what different CSP measurements are measuring and on how measurements need to be interpreted in estimation processes in order to derive meaningful implications. The next section is highlighting the research objectives and the main research question of this project. Thereafter, the research methods and brief summaries of each paper will be presented. Complementary, Section 5 is providing the main contributions and implications, while Section 6 is listing the research papers as attached in the annex.

[2] Here, innovation will solely be viewed from an input perspective and measured by Research & Development (R&D) intensity in line with the strategic management literature.

2. Research Objectives

This dissertation project is motivated by the fact that several measurements have been introduced to capture the multi-dimensionality of CSP. Although data availability has significantly advanced and various measurements can be considered valuable, also the corresponding weaknesses of different measurement approaches and its characteristics need to be evaluated. Accordingly, the following research question will be guiding the underlying analysis:

How relevant are different measurements of CSP
for the analysis of the relation between CSP and CFP?

Building upon this question, this project is aiming to address three complementary research objectives: First, disentangling influences resulting from CSP measurements on the CSP-CFP relationship by focusing on interferences with control variables, at the example of R&D intensity. Second, disentangling the underlying interrelations between improvements of corporate environmental data (e.g. CO_2 emissions) and environmental ratings provided by external rating agencies. Finally, analyzing the materiality of improvements in corporate environmental data.

2.1 Disentangling how CSP measurements influence the CSP-CFP relation focusing on interferences with control variables

The interrelations between CSP and R&D intensity, when examining the CSP-CFP link, were notably analyzed for the first time by McWilliams and Siegel (2000). McWilliams and Siegel (2000) have indicated that both constructs are measuring similar aspects to a certain extent. Thus, including R&D intensity in their research model led to a neutral relationship between CSP and CFP. Accordingly, this dissertation is aiming to analyze whether the frequency of ambiguous findings has been reduced in subsequent studies as predicted and whether different measurements of CSP are impacting the results. Utilizing significant data improvements and the emergence of further measurements for CSP, this thesis is also aiming to replicate and extend the original econometrical analysis to disentangle effects resulting from different CSP measurements.

2.2 Identify interrelations between corporate environmental data and environmental ratings

It is well established that CSP is a multidimensional construct (Wood 2010). Thus, highly aggregated CSP measurements have been used to account for this multidimensionality (e.g. Barnett and Salomon 2012; Callan and Thomas 2009; Griffin and Mahon 1997; McWilliams and Siegel 2000). Here, multiple CSP aspects are blended into aggregated variables to assess the firm's overall performance. Next to aggregated measurements, scholars increasingly employ more disaggregated operational measurements directly provided by the firms, such as CO_2 emissions (e.g. Dixon-Fowler et al. 2013; Mattingly and Berman 2006; Van der Laan et al. 2008). Both approaches are important to analyze internal and external views on CSP. Accordingly, this dissertation project is addressing the interrelations between these two different measurements to foster the understanding whether the improvements on an operational disaggregated level (e.g. CO_2 emissions) are reflected in aggregated measurements (e.g. CSP scores).

2.3 Assess materiality of improvements in corporate environmental data

Traditional views on CSP attempt to make sense of the positive relationship between levels of CSP and CFP. Based on a best-in-class view, companies with the highest level of CSP are also achieving the highest financial performance (ceteris paribus). However, this view fails to consider if improvements in CSP will also be rewarded. Thus, changes in CSP, expressed in either improvements or deteriorations, can also be expected to influence financial performance. From this view, companies that improve the most could be considered best-in-progress companies. This third and final paper project of the dissertation reassesses whether these continuous CSP improvements are financially material and provide incentives to further improve for leaders as well as for laggards.

Figure 1: Research focus of the individual papers

To address the research objectives, three complementary papers have been developed as highlighted by Figure 1. Paper I and II are addressing the first research objective by reviewing the findings of studies published after McWilliams and Siegel's (2000) indication of the omission of R&D intensity as well as by replicating and extending the original study, employing different CSP measurements. Paper III addresses research objectives two and three. Here, further disaggregated CSP measurements on the operational level are utilized to analyze how continuous improvements of operational disaggregated data are reflected within aggregated ratings provided by rating agencies as well as how these improvements are related to CFP.

3. Methods

This section covers the different methods utilized in the individual research papers. In general, this dissertation project is using quantitative research methods such as a systematic literature review, a vote count analysis, and econometrical analyses to study the underlying factors in the CSP-CFP relationship.

The first paper is based on a systematic literature review in order to analyze and synthesize recent empirical findings. Systematic literature reviews are prominently used in multiple sciences (e.g., medical science) in order to categorize existing knowledge on a specific topic and provide evidence for informing practice, policy making, and future research (Khan et al. 2003; Tranfield et al. 2003). Accordingly, a clearly defined review protocol has been developed in order to account for replicability and transparency. Here, research questions have been outlined to analyze how the indication of the misspecification issue by McWilliams and Siegel (2000) is reflected in subsequent studies. Based on these questions, inclusion and exclusion criteria have been defined in order to identify relevant studies.

Afterwards, quantitative aspects of the selected studies have been collected and assessed to allow for a vote count review. Based on the selected information, the vote count review has been customized to the individual research questions. Vote count reviews have been used in initial literature reviews in order to gain a first understanding of the underlying mechanisms (Griffin and Mahon 1997; Margolis and Walsh 2001). The empirical findings of different studies were categorized into positive, negative and neutral relationships between CSP and CFP depending on the statistical significances and directions. Nevertheless, vote count techniques also have been criticized as vote count assessments “do not correct for sampling and measurement error” (Orlitzky et al. 2003 p. 410). Accordingly, statistical errors are more likely to influence the results compared to more advanced meta-analyses. Nevertheless, studies have shown that specifically in this debate the results remained relative stable across different synthesizing approaches (Friede et al. 2015). Therefore, this combined approach of quantitative literature review and vote count analysis has been selected to foster the transparency how results have reflected the indication of the misspecification problem as suggested by McWilliams and Siegel (2000).

In the subsequent research paper projects, advanced econometrical approaches have been chosen to revisit specific issues in the CSP and CFP relationship. Accordingly, the second and third research papers employed panel regression analyses. Panel regression analyses are very useful when analyzing longitudinal data in order to examine underlying relationships between variables. In this statistical approach, independent variables are examined in order to identify potential relationships with a defined dependent variable over time and across observations. Here, specifically different measurements of CSP and R&D intensity are in focus as independent variables in order to examine their relationship with the dependent variable, CFP. Panel analyses allow to assess two types of variances, between- and within-firm variance (Certo et al. 2017). Accordingly, specification tests have been performed and suggested the use of firm and year fixed effects regression models.

Fixed effects estimations are useful in order to address one of the main concerns McWilliams and Siegel (2000) pointed out. By controlling for unobserved heterogeneity, fixed-effects estimations mitigate issues arising from omitted variables bias which results into endogeneity issues (Bettis et al. 2014). Nevertheless, endogeneity issues can also arise from other issues such as measurement errors and simultaneity (Surroca et al. 2010). Measurement errors occur when the true value of a variable cannot be measured directly. In this case, the measurement of a variable is a sum of its true value and an error term. Simultaneity refers to the situation where independent and dependent variable are interrelated (Wooldridge 2010). Here, CSP would not only be related with CFP, but CFP would also be related with CSP. Different strategies have been used in order to mitigate issues resulting from these endogeneity concerns. Accordingly, different measurement approaches, instrumental variable estimations and the introduction of lagged variables have been utilized in order to account for measurement errors and simultaneity (Bascle 2008; Zaefarian et al. 2017). Finally, to further foster the robustness of the analyses, the estimations have been replicated utilizing different data sources.

4. Summaries of the Papers

4.1 R&D – the missing link between Corporate Social Performance and Financial Performance?

McWilliams and Siegel's (2000) analysis marked a turning point in the empirical analysis of the relationship between CSP and CFP. By highlighting the omission of important variables such as R&D intensity, they identified that model misspecifications are one of the sources that have led to the ambiguous results, when analyzing the relationship between CSP and CFP. Controlling for R&D intensity, their results reveal that the underlying relationship between CSP and CFP, in fact, becomes neutral. Accordingly, this first paper investigates whether the introduction of R&D intensity as a control variable has been widely accepted and ultimately resulted into more neutral findings.

Based on a systematic literature review incorporating a quantitative vote count, 97 subsequent studies have been identified that employ comparable research models. The results suggest that a growing number of studies incorporated R&D intensity in their research models. However, contrary to the original assumptions, the inclusion of R&D intensity has yielded more findings with a positive relationship between CSP and CFP, while multiple different measurements of CSP have been used in the studies. In addition, R&D intensity tends to be significantly related to CFP, but also pointing in a negative direction. According to these findings, different options for methodological advancements have been derived, which form the basis for the subsequent two papers. In a first step, an enhanced replication study is recommendable in order to gradually analyze how the results are reflected within an updated sample. In a second step, this could be complemented by refining different measurement approaches of CSP in order to clearly disentangle the effects of aggregation and disaggregation and better understand how measurements drive the relationship. The results are adding to the literature by emphasizing that innovation in terms of R&D intensity and CSP are not per se tending to cover similar strategic aspects. Instead, further econometrical analyses are needed to revisit and disentangle issues resulting from omitted variables and measurement errors.

4.2 Corporate Social and Financial Performance: Revisiting the role of R&D

Motivated by the ambiguous findings of the systematic literature review (4.1), this paper is an enhanced replication of McWilliams and Siegel's (2000) original study. The original stipulations have not been consistently verified by the findings of subsequent studies. Thus, this study aims to disentangle the effects of R&D intensity and CSP in order to understand the effects resulting from different measurements of CSP.

Methodologically, three complementary econometrical steps have been applied in order to foster the transparency of the underlying relationships. First, guided by McWilliams and Siegel's (2000) approach, the same research model and measurements have been replicated using an updated dataset. Second, the measurement of CSP has been revised in accordance with findings from recent studies to account for construct validity and to utilize the panel structure (Mattingly and Berman 2006; Orlitzky et al. 2003). Here, highly aggregated CSP measurements like the affiliation with sustainability indices have been criticized as they solely result into binary measurements. Instead, KLD scores at different aggregation levels have been used to allow for more variation. Lastly, the functional relationship between R&D intensity and CFP has been analyzed, which was motivated by the fact that studies have reported positive as well as negative relationships between R&D intensity and CFP.

The results indicate that there is a neutral relationship between CSP and CFP, when controlling for R&D intensity, but only for highly aggregated CSP measurements. While analyzing more disaggregated CSP measurements, a positive relationship between CSP and CFP has been identified. Specifically, CSP concerns are negatively related to CFP. In other words, reducing CSP concerns would improve CFP. Moreover, the results indicate that the functional relationship between R&D intensity and CFP is curvilinear. Thus, following a u-shaped curve, only high or low levels of R&D intensity contribute to a firm's financial performance. These findings contribute to the literature by highlighting the impacts of different CSP measurements and corresponding model designs as well as by emphasizing that CSP and R&D intensity cover different aspects of a firm and need to be managed separately.

4.3 A change will do you good: Do continual environmental improvements matter?

The previous analyses have underlined the importance of how CSP[3] is measured. Previous studies focused on aggregated measures provided by rating agencies and analyzed how certain levels of performance are related with CFP. With growing transparency on operational environmental data, such as CO_2 emissions or water consumption, new paths are emerging to better grasp the individual aspects of CSP and to better understand how operational data is reflected in CSP ratings. In addition, this would also constitute a shift in our understanding how

[3] Paper III is mainly focusing on CSP in terms of corporate environmental performance (CEP). Analyzing continual improvements, environmental data has been identified to be more suitable due to the availability of higher levels of measurement.

mangers, decision makers and investors are perceiving CSP. From this view, one need to acknowledge that CSP goes beyond complying and implementing standards or norms, but is a continuous improvement activity of a firm’s environmental and social impacts.

Accordingly, this paper aims to foster the transparency regarding the interdependence and materiality of continuous CSP improvements. First, this study is analyzing how operational CSP improvements (e.g. energy usage, CO_2 emissions, water usage) are reflected in aggregated CSP scores (e.g. KLD scores). Second, the materiality of continuous environmental improvements is assessed by reviewing the relation between the same operational CSP improvements and CFP in terms of ROA and Tobin’s Q.

On the one side, the analyses reveal that improvements of environmental impacts are not captured by CSP scores compiled by rating agencies. On the other side, the results indicate that improvements contain financially material information. A positive significant relationship with CFP can be found for improvements in all environmental variables. The findings point out two important contributions. First, researchers and practitioners should focus on both aspects of CSP, not only higher values of aggregated CSP but also high improvements of disaggregated operational CSP, as both contain financially material information. Focusing on continuous improvements of impacts constitute a shift of focus as this is not only fostering best-in-class companies, but also best-in-progress companies. Second, we show that aggregated CSP ratings do not adequately capture all aspects of CSP. Most scores rather focus on processual aspects of CSP instead of operational impacts. However, specifically managing impacts becomes increasingly important to stakeholders (especially investors).

5. Discussion and Conclusion

5.1 Main research findings

Revisiting the debate on the relationship between CSP and CFP, this dissertation project is aiming to address the relevance of different CSP measurements. Consequentially, three complementary research objectives have been addressed by three research papers. Paper I and II are aiming to disentangle how CSP measurements influence the CSP-CFP relation focusing on interferences with control variables, at the example of R&D intensity. Paper III is aiming to assess whether disaggregated operational CSP is reflected within aggregated CSP data provided by rating agencies and whether disaggregated operational CSP data contains financially material information.

Figure 2 summarizes the main results of this dissertation project, which are discussed in the next sub-sections. First, by disentangling the relationships of CSP and R&D intensity the results indicate that highly aggregated CSP measurements tend to cover similar aspects like R&D intensity resulting into a neutral relationship between CSP and CFP. However, when further disaggregating CSP into strengths and concerns, a positive relationship between CSP and CFP can be discovered. Second, improvements of disaggregated operational CSP, e.g. CO_2 emission improvements, are not reflected in aggregated CSP ratings (e.g. KLD Environmental Score). Finally, improvements of disaggregated operational CSP contain financially material information as these improvements are positively associated with CFP.

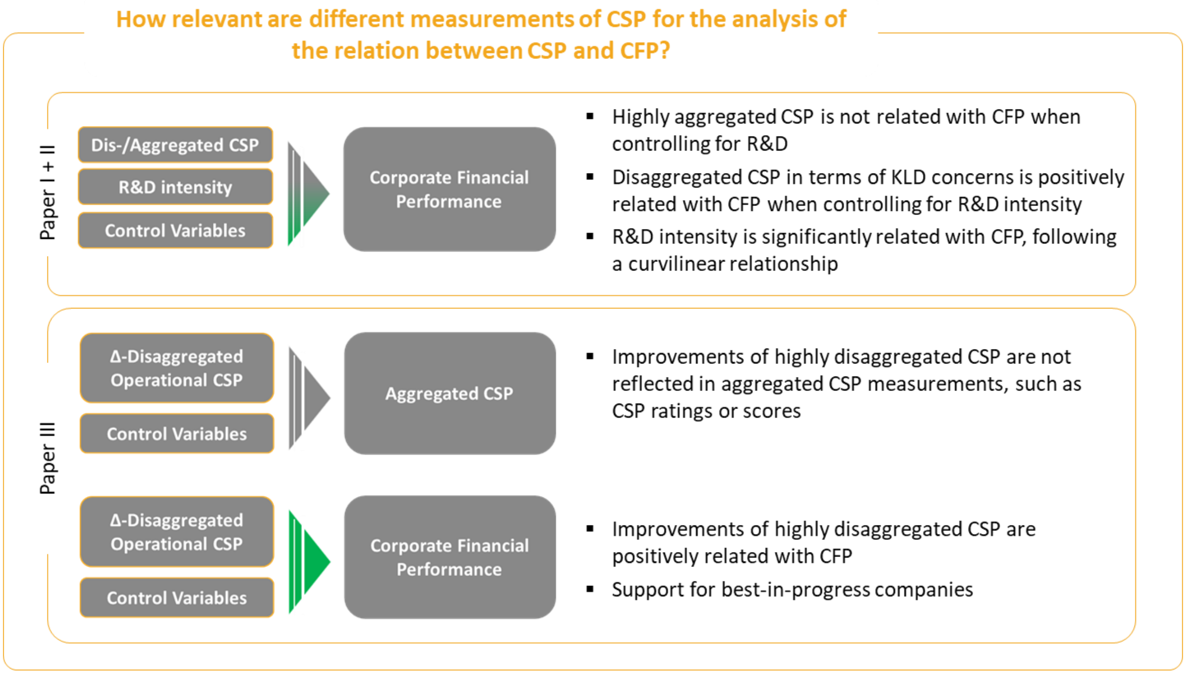

Figure 2: Overview of main research findings

5.1.1 Disentangling the relationships of CSP and R&D intensity

The publication of McWilliams and Siegel (2000) marked a significant turning point in this debate by indicating the omission of R&D intensity in research models. Reviewing subsequent studies, R&D intensity has been increasingly integrated in research models. However, the quantitative literature view conducted in Paper I shows that despite these efforts, researchers still report positive, negative and neutral relationships between CSP and CFP. The results indicate that CSP and R&D intensity do not ultimately tend to mimic similar strategic aspects, but the underlying mechanisms are more complex than expected. Different measurements of CSP have been identified as one key element that is differing between the studies. Accordingly, the underlying assumptions stipulated by McWilliams and Siegel (2000) need to be adjusted based on the employed measurement approach and understanding of CSP.

Building upon these findings, Paper II is revisiting McWilliams and Siegel's (2000) stipulations by utilizing advanced empirical estimation techniques and addressing issues arising from measurement errors due to highly aggregated CSP measurements. Mattingly and Berman (2006) have indicated that aggregation approaches have suffered from limited validity and reliability and need to be handled carefully in estimation processes. Paper II has contributed to these findings by replicating the original model analyzed by McWilliams and Siegel (2000) and extending their approach by using CSP measurements on different aggregation levels. The findings suggest that disaggregation will provide more transparency on the underlying mechanism. Specifically, the results show a positive relation between CSP and CFP in the case of CSP concerns (i.e. decreasing concerns will improve CFP).

While analyzing the different CSP measurements, R&D intensity remains very stable over the different research models. Here, R&D intensity is negatively associated with CFP, which seemed counterintuitive but still in line with selective previous studies (e.g. Barnett and Salomon 2012). This would imply that improving R&D intensity as a measurement for innovation would negatively impact CFP. Thus, building upon findings from the innovation management literature, the functional form between R&D intensity and CFP has been revised. Ultimately, a curvilinear relationship between R&D intensity and CFP has been revealed. Accordingly, only very low or high R&D intensities will be financially rewarded. Comparable findings are reported in the innovation literature by Bracker and Ramaya (2011) as well as Huang and Liu (2005). In contrast to the other studies, however, a U-shaped relationship has been reported instead of an inverted relationship. This contrary finding can be explained by the different statistical estimation approach. The application of firm fixed-effects estimation allows for individual analyses of R&D intensity on the firm level, while previous studies analyzed R&D intensity from an industry level. Consequentially, the results emphasized that R&D intensity and CSP need to be managed carefully and separately in order to improve the firm's financial performance.

5.1.2 Analyzing the relationship between improvements of disaggregated operational CSP and aggregated CSP ratings

Paper III analyzed operational environmental data and their relationship with aggregated CSP data provided by rating agencies. Here, the findings highlight that improvements of operational CSP are not reflected in aggregated CSP data. These results imply that CSP data provided by rating agencies only captures a snapshot of firm activities and is not able to reflect improvements in environmental impacts. Accordingly, both measurements tend to measure different aspects of CSP.

Previous studies pointed out that most of the CSP rating providers capture CSP with broad measurements that cover multiple dimensions, including both processes and impacts (Chen and Delmas 2011; Wood 2010). However, the underlying methodology for these CSP measurements is not standardized and widely confidential, which results in nontransparent methodologies. The lack of transparency has raised significant concerns with regards to construct validity and reliability. Many CSP measurements accumulate the results of "complex questionnaires and an analysis of public information sources" (Escrig-Olmedo et al. 2010, p. 445). We also know that many of these questionnaires are using dichotomous measures, which leads to binary data. Accordingly, CSP data is an accumulation of many qualitative factors to construct a CSP score (Carroll 2000; Chen and Delmas 2011). Disaggregated operational CSP data, on the other side, captures the individual inputs, outputs, and impacts of company's activities. Here, direct environmental inputs and outputs have been utilized to assess the firm's performance. This measurement approach is opening a more operational view on CSP as it is closely related the operations and environmental impacts. Yet, this view is not directly capturing the implementation and compliance with norms and standards as reviewed by external rating agencies, but the corresponding outcomes.

5.1.3 Analyzing the materiality of improvements of disaggregated operational CSP

As operational CSP is not reflected in aggregated CSP ratings, an additional analysis has been conducted to assess its financial materiality. The results strongly indicate that improvements in disaggregated operational CSP are positively associated with CFP. In other words, firms that reduce their environmental impacts are financially rewarded. These findings add to the existing literature by providing another ratio for companies to measure CSP, namely focusing on improvement ratios. In traditional analyses, the results indicate that companies with the highest CSP levels are rewarded the most (Brammer and Millington 2008; Lourenço et al. 2012; Orlitzky et al. 2003). This supports the paradigm of best-in-class companies. Here, the top performers in terms of CSP are rewarded. However, our results suggest that also best-in-progress companies will obtain financial benefits, which is reflected by internal (ROA) and external indicators (Tobin's Q). This provides additional insights in two ways. First, current CSP laggards are encouraged to improve their CSP as not only top performers, but also top improvers are already financially rewarded. Second, the top performers are also encouraged to continue their efforts and reduce their environmental impacts.

The results are considerably determined by the different measurement approaches. Disaggregated operational CSP data is mainly focused on outcomes, while aggregated CSP ratings are rather focused on processes or an aggregation of processes and outcomes (Wood 2010). Accordingly, both measurements cover different aspects of CSP and allow for different assessments. Specifically, disaggregated operational CSP data is suffering less from validity issues and allows for greater within-variation as we integrate continuous individual measurements (e.g. water usage, CO_2 emissions). Consequentially, both measurements tend to contain complementary financially material information, but need to be analyzed carefully to derive meaningful implications for academics, investors and managers.

5.2 Contribution

By disentangling the relationships of R&D intensity and CSP as well as analyzing the impacts of CSP measurements on the relationship between CSP and CFP, this dissertation project is contributing to the strategic management literature in two main regards. First, disentangling the effects of CSP and innovation, in terms of R&D intensity, Paper I and II are addressing an issue that has been a considerable turning point in the CSP-CFP debate (McWilliams and Siegel 2000). The findings add to the existing literature by pointing out how different measurements of CSP impact the original stipulations by McWilliams and Siegel (2000) as well as by revising the functional relationship between R&D intensity and CFP. Measurements of CSP have tremendously suffered from validity and reliability issues which not only impacted the CSP-CFP relation, but also the relations between the control variables and CFP. Accordingly, different approaches have been suggested to mitigate issues arising from measurement errors. Utilizing different measurements of CSP, our results confirm that R&D intensity is an important control variable, but that it covers different aspects than CSP. The results show that depending on the level of aggregation the relationship between CSP and CFP can be neutral or positive, while R&D tends to have a strong curvilinear relationship with CFP. These findings are important for academics, investors and managers as they indicate that measurement errors need to be considered carefully to derive meaningful implications. Here, we emphasize that CSP and R&D intensity need to be addressed carefully and separately by companies to be financially rewarded.

Second, analyzing the impacts of CSP measurements on the relationship between CSP and CFP, it has been recognized that there are two main ways to measure CSP. On the one hand, CSP has been traditionally measured on an aggregated level, e.g. as a score that incorporates adherence and compliance activities provided by third-party rating agencies. Consequentially, firms with the best ratings are expected to be rewarded the most. There is strong evidence for this hypothesis, however, it disregards the dynamic capabilities of CSP improvements. On the other hand, disaggregated measures of CSP provide indications of firm impacts on an operational level, which allows researchers and other stakeholders to investigate new phenomena, such as continuous improvements. By focusing on the improvements, the view is shifting from maintaining and aspiring certain levels of CSP and moving towards a focus on successive reductions of environmental impacts.

Accordingly, the findings of this dissertation project add also to the literature on the business case for sustainability by broadening the traditional view and complementing it with new knowledge, i.e. how improvements of disaggregated CSP are reflected within CSP ratings as well as by analyzing the materiality of CSP improvements. Surprisingly, the results show that CSP improvements are not captured in aggregated CSP ratings, but still contain financially material information. Thus, companies that improve the most on their operational CSP will also be financially rewarded. This directly relates back to the issue of measurement errors. CSP is a latent construct that has been measured in many ways. Utilizing the advancement of data, disaggregated operational data has become much more available in recent years. Thus, the findings are founded on these transparency improvements and adding a further component to better understand and analyze the available data.

Consequentially, these findings have three main implications for managers, investors, and scholars. For managers, the results imply that CSP is a complex construct that needs to be handled carefully in order to derive the right decisions. Multiple measurement approaches are available for CSP and they contain different information. Notably, disaggregation has indicated that ultimately CSP and R&D intensity need to be managed separately in order to minimize CSP concerns and ensure firm innovativeness. Additionally, CSP measurements contain different information. Complying with and implementing standards, norms and systems are mainly covered by CSP ratings. Ensuring compliance and integration will help to mitigate risks. However, CSP also has a dynamic perspective. These dynamics are only revealed when analyzing disaggregated operational CSP data. Operational data is continuously measured on a regular basis and capturing a firm's operational impacts. Both data sources contain financially material information, need to be viewed complementary and need to be managed by firms.

For investors, these results point out how careful CSP data needs to be handled to guide future decision making. CSP rating agencies provide one of the key sources for benchmarking CSP. However, the results emphasize that dynamic capabilities of operational impacts are not reflected. Given the growing number of impact investors, it is crucial to also integrate improvements of environmental footprints into evaluations, especially as they contain material information. In addition, new avenues are opening up for investors while the fostering of dynamic improvements does promote continuous efforts and not only reward best-in-class, but also best-in-progress companies.

For scholars, the results re-emphasize the importance of disentangling the different measurements of CSP. Many CSP measurements suffer from severe construct validity and reliability concerns that need to be carefully addressed in estimations. Our results highlight that the level of aggregation also significantly impacts the relationship between important control variables and CFP. Next to that, disaggregation also allows to capture the dynamic capabilities of CSP to understand the different aspects and in order to foster a holistic understanding of CSP.

5.3 Suggestions for Future Research

This dissertation provides three crucial suggestions for future research. First, additional research projects on this topic should not only focus on highly aggregated CSP measurements in their research designs. While aggregated CSP measurements try to provide a holistic view on CSP, this aggregation creates severe validity and reliability issues. Additionally, current measurements are also not able to integrate the dynamic capabilities of CSP. The analysis has shown that CSP needs to be viewed form two different perspectives. There are certain more stable CSP aspects that are usually well captured in aggregated CSP ratings, but also disaggregated operational CSP data is become rapidly available directly reported by the companies. Accordingly, the results rather encourage the development of new aggregated CSP data to account for the multi-dimensionality of CSP, but only to the extent where validity and reliability requirements are properly met. Here, operational CSP data provided by companies constitute a promising starting point to gather meaningful data and construct meaningful measurements.

Second, this dissertation project was largely focused on the operationalization and measurement of CSP and further control variables. This focus on the individual independent variables is of key importance to disentangle influences arising from measurement errors. However, the results encourage to broaden the view and re-assess correspondingly the dependent variables. Similar to CSP, CFP is also a multi-dimensional construct and multiple measurements have been proposed. Recent studies already integrate debt-market measurements, e.g. in order to better understand the relationship between CSP and bond spreads or ratings (La Rosa et al. 2018; Li et al. 2020; Salvi et al. 2020). Accordingly, the findings of this dissertation project could be utilized to further analyze different CFP measurements in order to better understand the impacts of CSP measurements.

Third, future research should focus on the different capacities of CSP. Traditionally, only best-in-class companies have been addressed by scholars. Best-in-class companies were expected to achieve higher financial performance and, in addition, be eligible to CSP indices. Accordingly, most CSP indices only consist of high performing CSP companies. However, our results indicate that we also need to focus on best-in-progress companies. In other words, also improvements of current CSP laggards will be financially rewarded. This leads to further implications for portfolio construction and the development of CSP indices. Therefore, further studies in Sustainable Finance would be very recommendable to analyze the dynamic capabilities of CSP on portfolio returns and associated risks.

6. Papers

The papers are included in the annex of this dissertation. The submission status of the papers is as of October 16th, 2020.

1. Schnippering, M. (2020). **R&D: the missing link between corporate social performance and financial performance?**, *Management Review Quarterly* 70(2), 243-255. [4]
2. Busch, T. and Schnippering, M. (2020). **Corporate Social and Financial Performance: Revisiting the role of R&D**, *Business & Society*, submitted.
3. Busch, T., Johnson, M., Schnippering, M. (2020). **A change will do you good: Do continual environmental improvements matter?**, *Strategic Management Journal*, submitted.[5]

[4] An earlier version of this paper has been presented at the 9th PRI Academic Network Conference 2017 in Berlin, Germany.

[5] Earlier versions of this paper have been presented at the GRONEN Research Conference 2018 in Almeria, Spain and at the 79th Annual Meeting of the Academy of Management 2019 (AOM) in Boston, Massachusetts, USA.

References

Allouche, J. & Laroche, P. (2005). A meta-analytical investigation of the relationship between corporate social and financial performance. *Revue de Gestion des Resources Humaines* 57 (July–September), 18–41.

Bansal, P. (2005). Evolving sustainably: A longitudinal study of corporate sustainable development. *Strategic Management Journal* 26(3), 197–218.

Barnett, M.L. & Salomon, R.M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal* 33(11), 1304–1320.

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* 6(3), 285–327.

Bettis, R., Gambardella, A., Helfat, C. & Mitchell, W. (2014). Quantitative empirical analysis in strategic management. *Strategic Management Journal* 35, 949–953.

Boulouta, I. (2013). Hidden connections: The link between board gender diversity and corporate social performance. *Journal of Business Ethics* 113(2), 185–197.

Brammer, S. & Millington, A. (2008). Does it pay to be different? An analysis of the relationship between corporate social and financial performance. *Strategic Management Journal* 29(12), 1325–1343.

Bracker, K. & Ramaya, K. (2011). Examining the impact of research and development expenditures on Tobin's Q. *Academy of Strategic Management Journal* 10, 63–79.

Callan, S.J. & Thomas, J.M. (2009). Corporate financial performance and corporate social performance. An update and reinvestigation. *Corporate Social Responsibility and Environmental Management* 16(2), 61–78.

Carroll, A.B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of Management Review* 4(4), 497–505.

Carroll, A.B. (2000). A commentary and an overview of key questions on corporate social performance measurement. *Business & Society* 39(4), 466–478.

Certo, S.T., Withers, M.C. & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within-and between-firm effects. *Strategic Management Journal* 38(7), 1536–1556.

Chatterji, A.K., Levine, D.I. & Toffel, M.W. (2007). Do corporate social responsibility ratings predict corporate social performance? Working Paper No. 33, Division of Research, Harvard Business School, Harvard University.

Chen, C.M. & Delmas, M. (2011). Measuring Corporate Social Performance. An Efficiency Perspective. *Production and Operations Management* 20(6), 789–804.

Dixon-Fowler, H. R., Slater, D.J., Johnson, J.L., Ellstrand, A.E. & Romi, A. M. (2013). Beyond “does it pay to be green?” A meta-analysis of moderators of the CEP–CFP relationship. *Journal of Business Ethics* 112(2), 353–366.

Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review* 20(1), 65–91.

Endrikat, J., Guenther, E. & Hoppe, H. (2014). Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal* 32(5), 735–751.

Escrig-Olmedo, E., Muñoz-Torres, M. J. & Fernandez-Izquierdo, M.A. (2010). Socially responsible investing: sustainability indices, ESG rating and information provider agencies. *International Journal of Sustainable Economy* 2(4), 442–461.

Friede, G., Busch, T. & Bassen, A. (2015). ESG and financial performance. Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment* 5(4), 210–233.

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

Griffin, J. J. (2000). Corporate Social Performance: Research Directions for the 21st Century. *Business & Society* 39(4), 479.

Griffin, J. J. & Mahon, J. F. (1997). The Corporate Social Performance and Corporate Financial Performance Debate. *Business & Society* 36(1), 5–31.

Huang, C. & Liu, C. (2005). Exploration for the relationship between innovation, IT and performance. *Journal of Intellectual Capital* 6(2), 237–252.

Jo, H., & Harjoto, M.A. (2011). Corporate governance and firm value: The impact of corporate social responsibility. *Journal of Business Ethics* 103(3), 351–383.

Jones, T. M. (1995). Instrumental stakeholder theory: A synthesis of ethics and economics. *Academy of Management Review* 20(2), 404–437.

Khan, K. S., Kunz, R., Kleijnen, J. & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine* 96(3), 118–121.

La Rosa, F., Liberatore, G., Mazzi, F. & Terzani, S. (2018). The impact of corporate social performance on the cost of debt and access to debt financing for listed European non-financial firms. *European Management Journal* 36(4), 519–529.

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care* 4(3), 324.

Li, Z., Tang, Y., Wu, J., Zhang, J. & Lv, Q. (2020). The interest costs of green bonds: Credit ratings, corporate social responsibility, and certification. *Emerging Markets Finance and Trade* 56(12), 2679–2692.

Lourenço, I.C., Branco, M.C., Curto, J.D. & Eugénio, T. (2012). How does the market value corporate sustainability performance? *Journal of Business Ethics* 108(4), 417–428.

Margolis J.D. & Walsh, J.P. (2003) Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly* 48(2), 268–305.

Margolis, J.D., Elfenbein, H.A. & Walsh, J.P. (2007) Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Working paper*.

Margolis, J. D., Elfenbein, H.A., & Walsh, J.P. (2009): Does it pay to be good… and does it matter? A meta-analysis of the relationship between corporate social and financial performance. *Working paper*, Harvard University, Cambridge, MA.

Margolis, J. D. & Walsh, J.P. (2001). People and profits?: The search for a link between a company's social and financial performance. *Psychology Press*.

Marti, C.P., Rovira-Val, M.R. & Drescher, L.G., (2015). Are firms that contribute to sustainable development better financially? *Corporate Social Responsibility and Environmental Management* 22(5), 305–319.

Mattingly, J.E. & Berman, S.L., (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society* 45(1), 20–46.

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

Orlitzky, M. & Benjamin, J.D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society* 40(4), 369–396.

Orlitzky, M., Schmidt, F.L., Rynes, S.L. (2003). Corporate social and financial performance: A meta-analysis. *Organization studies* 24(3), 403–441.

Peloza, J. (2009). The challenge of measuring financial impacts from investments in corporate social performance. *Journal of Management* 35(6), 1518–1541.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J. & Nykvist, B. (2009). Planetary boundaries: exploring the safe operating space for humanity. *Ecology and society* 14(2).

Salvi, A., Giakoumelou, A. & Bertinetti, G.S. (2020). CSR in the bond market: Pricing stakeholders and the moderating role of the institutional context. *Global Finance Journal*, 100522.

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., De Vries, W., De Wit, C.A. & Folke, C. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science*, 347(6223).

Surroca J. & Tribo, J.A. (2005). The Corporate social and financial performance relationship: What's the ultimate determinant? *Academy of Management* Proceedings, E1–E6.

Surroca J., Tribó, J.A. & Waddock, S. (2010). Corporate responsibility and financial performance. The role of intangible resources. *Strategic Management Journal* 31(5), 463–490.

Tranfield D., Denyer, D. & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14(3), 207–222.

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 US firms. *Academy of Management Review* 10(3), 540–557.

van der Laan, G., van Ees, H. & van Witteloostuijn, A. (2008). Corporate social and financial performance: An extended stakeholder theory, and empirical test with accounting measures. *Journal of Business Ethics* 79(3), 299–310.

Wood, D. J. (2010). Measuring corporate social performance: A review. *International Journal of Management Reviews* 12(1), 50–84.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. *MIT press*.

Zaefarian, G., Kadile, V., Henneberg, S.C. & Leischnig, A. (2017). Endogeneity bias in marketing research: Problem, causes and remedies. *Industrial Marketing Management* 65, 39–46.

Annex 1: First Research Paper

R&D – the missing link between Corporate Social Performance and Financial Performance?

Abstract

The relationship between corporate social performance (CSP) and financial performance (CFP) has been analyzed for decades. Despite these efforts, the results remain ambiguous. The omission of important variables in the econometrical estimation process is expected to be one reason for the mixed results. Accordingly, this study is focused on the role of R&D as pointed out by McWilliams and Siegel (2000). A quantitative literature and vote count review has been conducted in order to evaluate the acceptance, significance and influence of R&D as a control variable for analyzing the relationship between CSP and CFP. The results of this review are contrary to the initial assumptions stipulated by McWilliams and Siegel (2000). While a growing number of statistically significant R&D variables can be found in estimation models, the integration has rather led to an increase in findings with a positive CSP-CFP relation instead of findings with a neutral relation. These results have been used to derive implications for future research with regards to the materiality and operationalization of R&D.

Keywords: *R&D intensity; literature review; corporate social responsibility; firm performance*

1. Introduction

For decades, the relationship between corporate social performance (CSP)[6] and corporate financial performance (CFP) has been controversially discussed in academia (Ullmann 1985; Waddock and Graves 1997; Margolis et al. 2009; Barnett and Salomon 2012). Different empirical models have been specified and tested. However, to date there are still questions whether this relationship is positive, negative or neutral (Friede et al. 2015). McWilliams and Siegel (2000) have suggested one explanation for the ambiguous results. They argued that the controversy is founded on a misspecification of the underlying research models. They hypothesized that significant variables are omitted from the analysis and, therefore, the estimations are biased and inconsistent. Predominately, their analysis is focused on the omission of corporate innovativeness measured by R&D intensity. The findings led to the conclusion that the impact of CSP on CFP is neutral when R&D intensity is included in the model as it can be regarded as a kind of proxy for CSP. Accordingly, all previous estimation models that do not include a variable that accounts for R&D suffer from misspecification.

The research focus of this paper is directed towards the changes that have been induced by the article of McWilliams and Siegel (2000) and how the perception of R&D has developed. The indication of the misspecification issue could be a turning point in the CSP-CFP debate. Reviewing their stipulations is an important step in order to understand the relation between R&D intensity and CSP with regards to CFP. This triangular relationship could substantially influence the implications for managers and decision makers, especially if R&D intensity would have significant moderation effects on the CSP-CFP link. Thus, this review shall help to understand how the indication has affected the empirical results in this debate. In addition, this would foster the understanding, if CSP and R&D intensity cover similar aspects as proposed by McWilliams and Siegel (2000) or need to be handled separately.

[6] Here, CSP will be used as an umbrella term that incorporates social and environmental aspects.

Consequentially, the following analysis aims to provide a quantitative literature review which focuses on the influence of R&D intensity concerning the CSP-CFP debate since the indication of the misspecification problem. To guide this analysis the following research questions have been developed:

1. Did the empirical literature on the CSP-CFP relationship incorporate McWilliams and Siegel's suggestion to include R&D as a control variable?
2. Does the inclusion of R&D in fact yield neutral results regarding the CSP-CFP relation as predicted by McWilliams and Siegel?
3. Is the impact of R&D on the regression model statistically significant and does the sign of the coefficient point in a positive direction as predicted by McWilliams and Siegel?

To analyze these questions, the next chapters will provide an overview on important landmarks in this debate as well as the research methodology. In the fourth section, we present the results in accordance with the research questions, identifying a distress in the original assumptions and the actual empirical results based on our review sample. Finally, we derive options for methodological enhancements and conclude with limitations as well as implications for future research and practice.

2. The business case for sustainability: empirical landmarks

Analyzing the link between CSP and CFP has become increasingly important as a significant positive relation could justify the business case for sustainability (Schreck 2011). In this realm, CSP measures are valued as aspects that improve corporate financial performance and overall "societal expectations" (Waddock and Graves 1997, p. 304). Hence, a company should pursue sustainability measures to obtain competitive advantages (Orlitzky et al. 2003; López et al. 2007; Busch and Hoffmann 2011). On the other side, there is a paradigm that highlights the costs of sustainability measures. From this point of view, investing in CSP will not improve financial performance, but rather lead to additional costs that reduce financial performance (Friedman 1970). Accordingly, a company should only invest in CSP measures as far as it is obliged to, otherwise it would destroy value (McWilliams and Siegel 2000).

Based on these different views an extensive literature has evolved that rather led to a continuum between the two views than to a clear solution. Different quantitative and qualitative research methodologies have been employed. Basically, studies are focused on the validation of the relation between CSP and CFP (Günther et al. 2012; Endrikat et al. 2014). Therefore, empirical analyses remain at the core of the debate, but to date fail to present a clear piece of evidence for one perspective (Surroca and Tribo 2005; Barnett and Salomon 2012). These results are categorized in the three categories: positive, negative and neutral relationship (Aras et al. 2010; Günther et al. 2012). Increasingly, research is focused on methodological problems of the empirical analyses to explain the differences in the results (Elsayed and Paton 2005).

The basic econometric models regress the impact of CSP on CFP. Both variables are considered to be rather broad meta-constructs (Margolis and Walsh 2003; Friede et al. 2015). Various operationalizations have been applied to measure them. CSP variables are generally categorized into individual and aggregated CSP measures (Makni et al. 2009). In that regard, social as well as environmental dimensions or indices combined out of both are used to describe CSP

(Orlitzky et al. 2003; Surroca and Tribo 2005; López et al. 2007; Alvarez 2012). Again, also these constructs are rather broad and there is no standardized way available that defines clear indicators for social and environmental aspects (Zafar and Sulaiman 2019). Thus, scholars have used many different proxy variables like reputation indices, philanthropy measures or scores provided by sustainability analysts (Fauzi 2009). A growing focus can be seen concerning individual CSP measures in order to assess the specific aspects separately. These analyses are important in order to understand differences between social or environmental CSP measures and their impact on CFP (Mahoney and Roberts 2007; Schreck 2011). However, this multidimensionality increases the complexity of the debate and hampers the comparability as different studies use different constructs to measure CSP.

The measurement of CFP seems to be more straightforward, as the financial performance of corporations has been assessed for decades. However, a similar problem has developed. There is no clear agreement on the right measure for CFP. Mainly, the measurement approaches are divided into accounting and market-based views (Aras et al. 2010; Marti et al. 2015). Typical accounting-based measures include return on assets and return on equity (Günther et al. 2012). The drawback of this approach is its retrospective nature and inconsistency as it is based on accounting principles (Aras et al. 2010). Therefore, measurements via market-based variables have become more important. Tobin's Q, Market Value Added or development of stock prices are commonly used indicators for these approaches (Margolis et al. 2009). They are considered as rather forward-looking by integrating the shareholder expectations into the debate (Aras et al. 2010).

Additionally, control variables are used to improve the model fit (Garcia-Castro et al. 2010; Andersen and Dejoy 2011). Traditionally, the specification is founded on a close connection to findings from financial econometrics. Based on that, initial models have employed variables to control for risk, size and industry (Ullmann 1985). This set up has often been replicated

(Margolis et al. 2009; Alvarez 2012). However, over time more studies challenged the fitness of these control variables as model misspecifications could be the reason for the heterogeneous results (Surroca and Tribo 2005). Accordingly, more variables have been introduced and tested. Especially, intangible firm resources like R&D for innovativeness or human capital for organizational resources have been increasingly analyzed (McWilliams and Siegel 2000; Surroca and Tribo 2005; Ozkan 2018). In accordance, previous reviews have summarized the relationship between CSP and CFP, while analyzing the role of control variables (Margolis and Walsh 2003; Friede et al. 2015). Van Beurden and Gössling (2008, p. 411) highlighted the relevance of these “moderating variables” and their impact on the significance of the CSP-CFP relation. However, these reviews encounter the impact of the R&D variable from a rather broad point of view based on a limited number of studies. Therefore, we will provide a quantitative literature review focused specifically on the influence of R&D intensity concerning the CSP-CFP debate.

3. Methodology

Based on the research questions outlined before, a quantitative literature review has been conducted with a specific focus on quantitative findings in the literature in order to analyze the relationship between R&D, CSP and CFP (Fisch and Block 2018, p. 105).

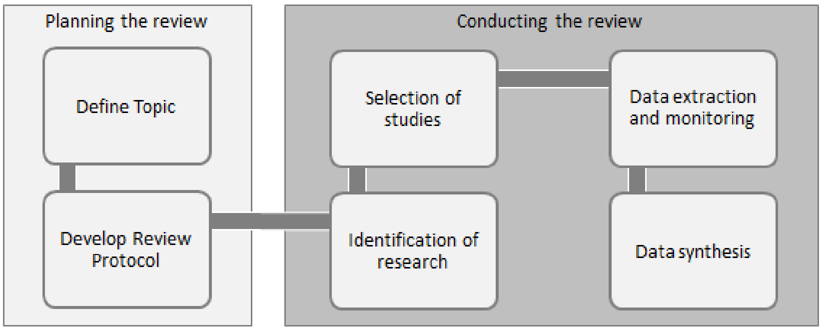

Illustration 1: Review Process, own illustration, based on Tranfield (2003).

Tranfield (2003) provided a practical process for systematic literature reviews as outlined in illustration 1. Originally, the research questions and the focus of this article have been developed as already outlined before. Based on this framework a research protocol has been created that defines the search strategy, initial screening, advanced inclusion and exclusion criteria as well as a draft for the database model.

The search strategy can be divided into two parts: computerized search and manual search. For manual search, reference lists from relevant primary studies have been searched as well as journals for further articles. Additionally, direct correspondence with other researchers has led to an enhanced database. For the computerized search, predefined keywords combined with Boolean algebra have been used to find material articles in a suitable search engine. ABI/INFORM by ProQuest has been used as the main search engine as it is the most complete database available and covers nearly all relevant business and management journals (McWilliams and Siegel 2000). Through predefined keywords, ABI/INFORM provides a detailed overview of relevant articles. The following keywords and syntax have been used to identify material publications:

(("Financial performance" AND ("Research and Development" OR R&D) AND (regression OR quantitative) AND ("Corporate Social Responsibility" OR CSR) OR ("Corporate Social Performance" OR CSP) OR (Environmental Social Governance OR ESG)) AND peer(yes)) AND la.exact("ENG")

General parameters have been predefined for both search strategies in order to secure the quality of the reviewed articles. Therefore, all articles needed to be in English, published in peer-reviewed journals and be published after 2000, coherent with the time McWilliams and Siegel published their article. Further, advanced evaluation criteria have been developed to manage the complexity and extent of articles. First of all, screening criteria have been developed as a filter to optimize the number of primary results. In that regard, titles and abstracts have been reviewed whether they are located within the field of CSP/CFP relationship analysis and employ quantitative statistical approaches. This step aims to improve the quality of the included studies regarding the underlying research focus. Afterwards, a set of exclusion criteria has been created for a more profound analysis of the remaining articles. Accordingly, studies have been excluded that synthesize results (either in meta-analysis or literature reviews), do not use regression models (e.g. excluding factor analyses, structured equation models), estimate solely curvilinear models and do not follow the causality that CSP has an impact on CFP.

The literature review has been conducted based on the outlined review protocol. The initial keyword-based search request at ABI/INFORM has led to 1,860 results. Additionally, 70 articles have been identified through manual search. In the first part of the selection process, these articles have been screened and unsuited publications have been excluded so that the number of results decreased to 272 articles. Out of these articles, 97 publications have been identified for the final database after excluding further studies based on their econometrical fitness. Two researchers have conducted the whole process in order to account for complementary validation (Tranfield 2003).

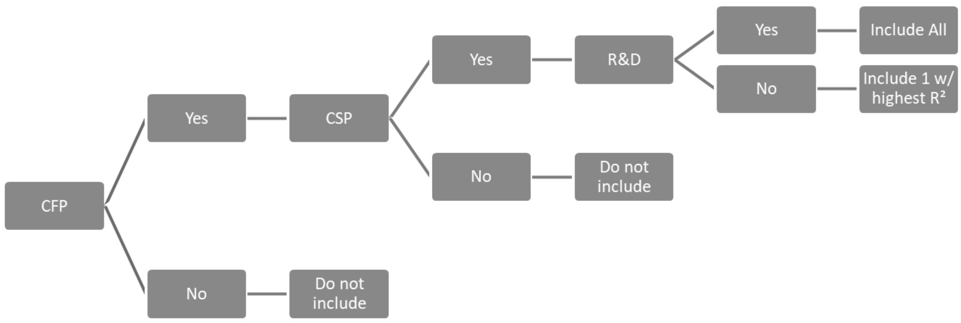

Illustration 2: Model selection process.

The conduction phase is complemented by data extraction and synthesis. Parameters have been defined in order to systematically extract information on the model design and regression results from the selected articles. Based on the multidimensional character of this debate, most of the studies use more than one regression estimation model to test different specifications. However, not all models are suitable to be integrated into this review as the main focus is directed towards the influence of R&D. Accordingly, as outlined by illustration 2, a decision tree has been developed in order to structure the selection of relevant models to identify representative models. First, all models with CFP measures have been selected. Afterwards, availability of CSP variables has been analyzed. Models without any CSP measure have been excluded. In the next stage, the use of R&D has been evaluated. All models that included a measure for R&D were included in the analysis and for models that have not accounted for R&D the model with the highest R^2 has been selected. Based on this selection process a total of 197 research models have been derived from the articles.

The extracted models have been analyzed and synthesized via general descriptive methods as well as by the vote-counting method. The vote counting technique has been often used by scholars to synthesize the findings in this debate (Lin et al. 2009). Hereby, the results of the individual models are counted and structured into three categories. These categories describe the relationship between CSP and CFP as positive, negative or neutral (Elsayed and Paton 2005). This basic method is disputed as it does not "correct for sampling and measurement errors" (Orlitzky et al. 2003, p. 410). Accordingly, the results provide weaker statistical evidence and lead to slightly biased results compared to more detailed meta-analyses (Wang et al. 2015). However, Elsayed and Paton (2005) as well as Friede et al. (2015) have not found any significant differences between the results of the vote counting compared to meta-analyses. Based on the aim of this review, the vote-counting has its usefulness in order to discover new areas for further research in a systematic way. Therefore, it should be rather considered as a well substantiated starting point for an in-depth discussion.

4. Results

Based on the outlined review process 97 publications from 2000 to 2016 have been analyzed. Despite some exceptions, the number of publications has increased recognizable until reaching its peak in 2011 with 12 publications as highlighted in Figure 1. Since then the amount of publications has averaged out on a lower level, but gaining more prominence in 2016. This highlights the further importance of this debate in the academic discourse.

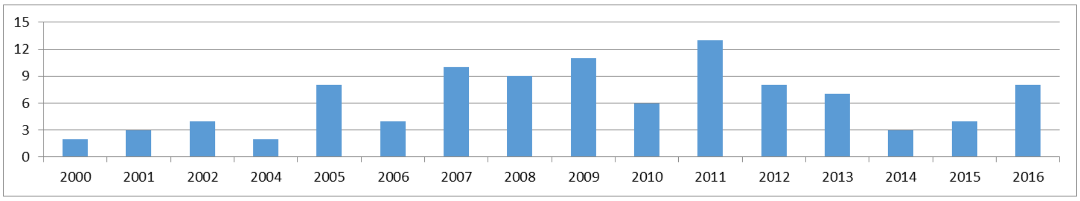

Figure 1: Number of publications per year from 2000-2016 (n = 97).

Out of these publications, 197 models have been extracted and reviewed. On average, around 60% of these models do not include R&D as a control variable while 40% consider R&D in the specification.

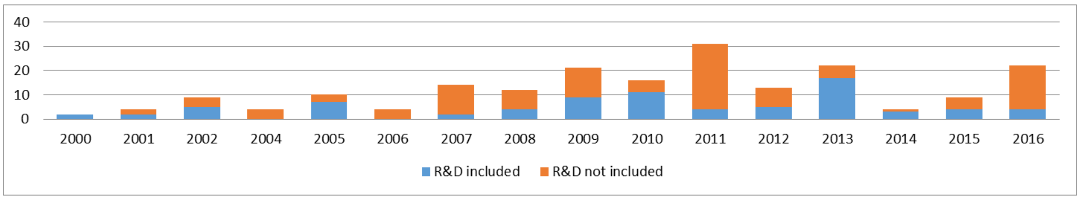

Figure 2: Integration of R&D in research models per year from 2000-2016 (n = 197).

The distribution of models that include R&D and models that exclude R&D does not appear to follow a specific trend as highlighted by Figure 2. From 2000 until 2005 relatively more models included R&D as a control variable. Hence, the integration of R&D has been quite prominent shortly after McWilliams and Siegel (2000) published their results. However, since 2006 on average more models have not integrated R&D variables. Especially in 2011 and 2016, almost all models do not consider R&D. In sum, the use of R&D has increased over the years indicating that R&D has become a considerable part of academic discourse. Nevertheless, there is still a noticeable amount of studies that do not account for R&D. One possible explanation could be related to data transparency. R&D information is not available for many companies outside of the US. As recent studies are more focused on analyses outside of the US, an inclusion of R&D would lead to a significant decrease in sample sizes, which shall be avoided in statistical estimations (Delmas et al. 2015).

4.1 Impact of R&D on the relationship between CSP and CFP

An assessment of the relationship between CSP and CFP as proposed by the second research question is founded on the general understanding that the estimation results of the individual models have been divided into three categories:

1. *Models that find a positive relationship between CSP and CFP*
2. *Models that find a neutral relationship between CSP and CFP*
3. *Models that find a negative relationship between CSP and CFP.*

The evaluation is based on the significance and direction of the CSP variable. A non-significant variable displays a neutral relationship and a significant positive or negative CSP variable vice versa the other categories. Additionally, another categorization has been developed to highlight the differences between the full sample and the sub-samples, where either R&D is included or excluded from the research model.

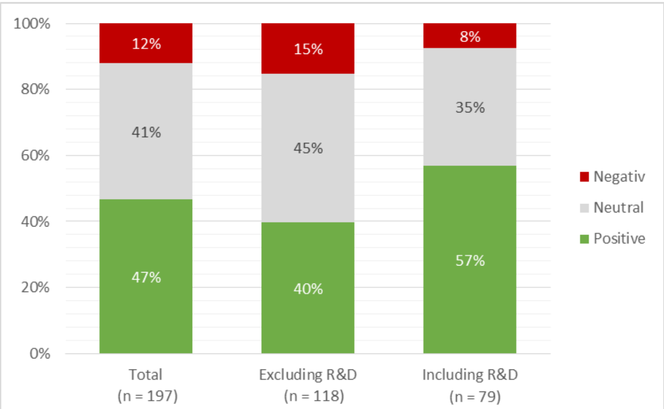

Figure 3: CSP and CFP relation in dependence of R&D.

When looking at the full sample, the relationship between CSP and CFP is consistent with other findings (Friede et al. 2015). 47% of the models indicate a positive relation between CSP and CFP. Further 41% account for a neutral relationship and the remaining 12% find a negative relationship between corporate social and financial performance. Hence, in general the association between CSP and CFP does not seem to destroy financial value.

Figure 3 highlights the changes in the distribution when R&D is included or excluded. For models where R&D is excluded from the set of control variables, only 40% of the statistical estimations find a positive relationship. Hence, omitting R&D leads to a decrease in positive findings. On the other side, there are slightly more findings that indicate neutral (45%) as well as negative (15%) relationships.

For models where R&D is included, the development is contrary. The integration of R&D leads to an increase in models that find a positive relationship (57%). Accordingly, the amount of neutral (35%) and negative (8%) findings decreases recognizably. These results are different from the prediction by McWilliams and Siegel (2000). According to their propositions, the relationship between CSP and CFP should be neutral when R&D is included in the model. However, the results of this review suggest a different development as the inclusion of R&D leads to a decrease in neutral relationships. Hence, the correlation between R&D and CSP could be depended upon additional factors.

4.2 Significance and direction of R&D

Based on the impact on the relationship between CSP and CFP, the models that include R&D (n = 79) have been analyzed more detailed in order to evaluate its robustness as indicated by the third research question. Therefore, in a first step, the significance of the R&D variable within these models has been analyzed. If a variable is considered as being significant, it is estimated to be different from zero and to add explanatory value to the model. Approximately, two third of the R&D variables are significant variables, while one third are non-significant. Hence, the inclusion of R&D seems to be useful based on the overall findings. Second, the directions of the significant R&D variables have been analyzed. Only significant variables have been reviewed as the non-significant variables have no explanatory impact on the model.

Intuitively, a positive sign for the coefficient of the R&D variable would be reasonable. Consequentially, an increase in R&D expenditures would be expected to lead to an increase in competitive advantage and, therefore, to an increase in financial performance (McWilliams and Siegel 2000). Hence, negative results would be rather counter-intuitive. Here, roughly 60% of the significant R&D variables are positive. Thus, the findings are not fully consistent with the theoretical expectations and the conflicting results should be carefully investigated in further analyses.

5. Conclusion

Decades of research have substantiated that the relationship between CSP and CFP is indeed complex. Over the years, different additional aspects have been introduced into this debate in order to improve the transparency on this relationship. Guided by three complementary research questions, this systematic literature review is evaluating the role of R&D in order to review the propositions made by Siegel and McWilliams (2000) and identify potential connection points for further research.

Did the empirical literature on the CSP-CFP relationship incorporate McWilliams and Siegel's suggestion to include R&D as a control variable?

The results indicate mixed findings with regards to the model specification. Directly after the indication of the relevance of R&D by McWilliams and Siegel (2000), most of the publications included R&D in their research models. Overall years, around two-thirds of the studies include R&D. Hence, it can be viewed as an important control variable. However, there is still a noticeable amount of models that do not consider R&D in their research design. Additionally, it seems as if recently fewer models account for R&D. Consequentially, the model specification did change, but the change process is ongoing as the influence of R&D is still difficult to grasp and data availability is further emerging.

Does the inclusion of R&D in fact yield neutral results regarding the CSP-CFP relation as predicted by McWilliams and Siegel?

It has been hypothesized that the inclusion of R&D would lead to a neutral relationship between CSP and CFP as found by McWilliams and Siegel (2000). However, this analysis has shown that the integration of R&D rather leads to an increase in positive relationships and a noticeable decrease in neutral results; i.e. contrary to the assumptions. These findings become even stronger for models where R&D is one of the main significant variables. A small increase in the neutral relationship could only be obtained from models where R&D is excluded from the set of control variables. Consequentially, R&D leads to different results with regards to this debate, but in favor of a positive CSP-CFP relationship.

Is the impact of R&D on the regression model statistically significant and does the sign of the coefficient point in a positive direction as predicted by McWilliams and Siegel?

The important role of R&D is founded on the assumption that R&D has a significant positive impact on the corporate financial performance. The results of this review highlight that most of the R&D variables are significant as well as positive. Hence, in accordance with the general assumptions, R&D can be considered as a valuable factor to add explanatory power to the estimation model. However, about 40 percent of the models show ambiguous results. Here, more evidence is needed regarding the operationalization of R&D. Some scholars use R&D intensities to account for size effects, but use different size deflators (e.g. sales, employees), while other models include pure R&D costs.

Based on these findings, multiple options for methodological enhancements become apparent in order to improve transparency with regards to the role of R&D. First, an enhanced meta-analysis might be suitable in order to foster the robustness of the results of this review. Differences in sample sizes or effect sizes are not specifically considered under the current setup as it solely differentiates between three distinctive categories (Elsayed and Paton 2005; Lin et al. 2009). Accordingly, scholars advocate to conduct a more precise meta-analysis to synthesize the results. Besides the correction for statistical artifacts, meta-analysis will also allow evaluating the role of moderator variables more precisely (Orlitzky et al. 2003). Endrikat et al. (2014) followed a similar approach in order to assess the moderating effects of the control variables. However, according to their set-up R&D was rather one of many addressed issues and the changes that result from R&D have not been investigated in detail. Further, the difference between models that include and exclude R&D has not been regarded as being significant although it almost reached the 5% acceptance interval (p-value = 6.5%). Further, methodological aspects regarding the operationalization of the R&D and CSP constructs might be re-evaluated in an enhanced replication study in order to understand the impact of these factors. Especially, the operationalization of R&D intensity is of key importance as the selection of the size deflator (e.g. sales) might influence the estimation results. Many scholars use the same variable set-up within their estimation models (Barnett and Salomon 2012; Garcia-Castro et al. 2010). Surroca et al. (2010) as well as Andersen and Dejoy (2011) are using slightly different approaches in order to include R&D. Surroca et al. (2010) divide R&D expenditures by the number of employees. Andersen and Dejoy (2011) do not use a ratio, but the total expenditures. However, all of these constructs are measuring innovation from an input perspective. On the other side, output variables (e.g. number of patents) could also be integrated in order to verify the consistency of the results. For CSP, similar problems arise. Using highly aggregated CSP variables has been widely criticized by scholars (Mattingly and Berman 2006). Especially in extreme cases, when multiple information is dichotomized into one binary variable (e.g. DSI 400), effect sizes for the underlying relationship are likely to be underestimated (Dawson and Weiss, 2002; MacCallum et al., 2002). Thus, McWilliams and Siegel's (2000) results could be influenced by the variable design.

As highlighted by McWilliams and Siegel (2000), the relationship between CSP and CFP is very complex, and the model specification is one of the most important aspects to avoid endogeneity bias. However, reviewing their stipulations with the recently published literature, our results indicate that especially two further empirical phenomena have emerged from their approach. First, the heterogeneity in the relationship between CSP and CFP has not eased through the inclusion of R&D. Instead of producing more neutral results, the inclusion of R&D rather leads to more positive results. And finally, R&D intensity seems to be significantly related to CFP, but frequently pointing in a negative direction. Thus, counterintuitively and contrary to McWilliams and Siegel (2000), investments in R&D are rather regarded as costs with a negative relation to CFP. For future research, the results imply that R&D is an important control variable and should be included in corresponding research models, but the inclusion does not necessarily lead to a neutral relationship between CSP-CFP. Instead, the relationship with CSP and CFP needs to be handled carefully in the estimation process and different specifications as well as operationalizations should be further investigated in order to avoid the introduction of further misspecification (e.g. negative relationship between R&D intensity and CFP). Decision makers, on the other hand, should acknowledge that R&D and CSP need to be specifically managed and that being innovative will not automatically influence a firm's CSP.

Bibliography

Alvarez IG (2012) Impact of CO 2 Emission Variation on Firm Performance. In: Bus Strat Env 21(7):435–454. DOI: 10.1002/bse.1729.

Andersen ML, Dejoy J (2011) Corporate social and financial performance. The role of size, industry, risk, R&D and advertising expenses as control variables. In: Bus Soc Rev 116(2):237–256. DOI: 10.1111/j.1467-8594.2011.00384.x.

Aras G, Aybars, A, Kutlu O (2010) Managing corporate performance. In: Int J Productivity & Perf Mgmt 59(3):229–254. DOI: 10.1108/17410401011023573.

Barnett ML, Salomon RM (2012) Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. In: Strateg Manag J 33(11):1304–1320. DOI: 10.1002/smj.1980.

Busch T, Hoffmann VH (2011) How Hot Is Your Bottom Line? Linking Carbon and Financial Performance. In: Bus Soc 50(2):233–265. DOI: 10.1177/0007650311398780.

Dawson NV, Weiss R (2012) Dichotomizing continuous variables in statistical analysis: A practice to avoid. Med Decis Making 32(2): 225-226.

Elsayed K, Paton D (2005) The impact of environmental performance on firm performance. Static and dynamic panel data evidence. In: Struct Chang Econ Dyn 16(3):395–412. DOI: 10.1016/j.strueco.2004.04.004.

Endrikat, J, Guenther E, Hoppe H (2014) Making sense of conflicting empirical findings. A meta-analytic review of the relationship between corporate environmental and financial performance. In: Eur Manag J 32(5):735–751. DOI: 10.1016/j.emj.2013.12.004.

Fauzi H (2009) Corporate social and financial performance: Empirical evidence from American companies. In: Globs Manag J 3:25-34.

Fisch C, Block J (2018) Six tips for your (systematic) literature reviewin business andmanagement research. Manag Rev Q 68:103–106.

Friede, G, Busch T, Bassen A (2015) ESG and financial performance. Aggregated evidence from more than 2000 empirical studies. In: J Sust Fin Inv 5(4):210–233. DOI: 10.1080/20430795.2015.1118917.

Friedman M (1970) The social responsibility of business is to increase its profits. In: The New York Times Magazine.

Garcia-Castro R, Ariño, MA, Canela, MA (2010) Does Social Performance Really Lead to Financial Performance? Accounting for Endogeneity. In: J Bus Ethics 92(1):107–126.

Günther E, Hoppe H, Endrikat J (2012) Corporate financial performance and corporate environmental performance: A perfect match? In: ZfU 34:279-296.

Lin CH, Yang HL, Liou DY (2009) The impact of corporate social responsibility on financial performance. Evidence from business in Taiwan. In: Technol Soc 31(1):56–63. DOI: 10.1016/j.techsoc.2008.10.004.

López MV, Garcia A, Rodriguez L (2007) Sustainable Development and Corporate Performance. A Study Based on the Dow Jones Sustainability Index. In: J Bus Ethics 75(3):285–300. DOI: 10.1007/s10551-006-9253-8.

MacCallum RC, Zhang S, Preacher KJ, Rucker DD (2002) On the practice of dichotomization of quantitative variables. Psychol Methods 7(1): 19-40.

Mahoney L, Roberts RW (2007) Corporate social performance, financial performance and institutional ownership in Canadian firms. In: Account For 31(3):233–253. DOI: 10.1016/j.accfor.2007.05.001.

Makni R, Francoeur C, Bellavance F (2009) Causality Between Corporate Social Performance and Financial Performance: Evidence from Canadian Firms. In: J Bus Ethics 89(3):409–422. DOI: 10.1007/s10551-008-0007-7.

Margolis JD, Walsh JP (2003) Misery loves companies: Rethinking social initiatives by business. In: Admin Sci Quart 48(2):268–305.

Margolis JD, Elfenbein HA, Walsh JP (2009) Does it pay to be good… and does it matter? A meta-analysis of the relationship between corporate social and financial performance. Working paper, Harvard University, Cambridge, MA.

Marti CP, Rovira VMR, Drescher LGJ (2015) Are Firms that Contribute to Sustainable Development Better Financially? In: Corp Soc Resp Env Ma 22(5):305–319. DOI: 10.1002/csr.1347.

Mattingly JE, Berman SL (2006) Measurement of Corporate Social Action. In: Bus Soc 45(1): 20–46.

McWilliams A, Siegel D (2000) Corporate social responsibility and financial performance: correlation or misspecification? In: Strateg Manag J 21(5):603–609.

Orlitzky M, Schmidt FL; Rynes SL (2003) Corporate social and financial performance: A meta-analysis. In: Organ Stud 24(3):403–441. DOI: 10.1177%2F0170840603024003910

Ozkan, N (2018) Replicating the R&D investments and financial structure relationship: evidence from Borsa İstanbul. In: Manag Rev Q 68(4): 399-411.

Saeidi SP, Sofian S, Saeidi P, Saeidi SP, Saaeidi SA (2015) How does corporate social responsibility contribute to firm financial performance? The mediating role of competitive advantage, reputation, and customer satisfaction. In: J Bus Res 68(2):341–350. DOI: 10.1016/j.jbusres.2014.06.024.

Schreck P (2011) Reviewing the Business Case for Corporate Social Responsibility: New Evidence and Analysis. In: J Bus Ethics 103(2):167–188. DOI: 10.1007/s10551-011-0867-0.

Surroca J, Tribo JA (2005) The Corporate social and financial performance relationship: What's the ultimate determinant? In: Acad Manag Annual Meeting Proceedings E1-E6. DOI: 10.5465/AMBPP.2005.18783433.

Surroca J, Tribó JA, Waddock S (2010) Corporate responsibility and financial performance. The role of intangible resources. In: Strateg Manag J 31(5):463–490. DOI: 10.1002/smj.820.

Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. In: Brit J Manage 14(3):207–222. DOI:10.1111/1467-8551.00375.

Ullmann, AA (1985) Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of US firms. In: Acad Manag Rev 10(3):540–557.

Van Beurden, P., & Gössling, T. (2008). The worth of values–a literature review on the relation between corporate social and financial performance. Journal of business ethics, 82(2):407-424.

Waddock SA, Graves SB (1997) The corporate social performance-financial performance link. In: Strateg Manag J 18(4):303–319.

Wang DHM, Chen PH, Yu THK, Hsiao CY (2015) The effects of corporate social responsibility on brand equity and firm performance. In: J Bus Res 68(11):2232–2236. DOI: 10.1016/j.jbusres.2015.06.003.

Zafar MB, Sulaiman AA (2019) Corporate social responsibility and Islamic banks: a systematic literature review. In: Manag Rev Q 69(2):159-207. DOI: 10.1007/s11301-018-0150-x.

Annex 2: Second Research Paper

Corporate Social and Financial Performance: Revisiting the role of R&D

Abstract: *Highlighting the role of R&D intensity, McWilliams and Siegel (2000) point out that the ambiguous results in the debate over corporate social performance (CSP) and financial performance (CFP) are due to a misspecification. However, two additional empirical phenomena have emerged. First, while controlling for R&D intensity, subsequent studies have continued to report ambiguous results and, second, many studies have found R&D intensity to be negatively related to CFP. Our study highlights two important insights. First, to avoid measurement errors, scholars should refrain from using highly aggregated CSP scores. For aggregated CSP, R&D intensity tends to mimic similar aspects leading to a neutral relationship between CSP-CFP, while disaggregated CSP reveals that CSP concerns are negatively associated with CFP and cannot be mitigated by improved R&D intensities. Second, our study reveals that controlling for and managing R&D is indeed essential, whereas the functional form of R&D intensity is not linear, but curvilinear.*

Keywords: *corporate social performance (CSP), corporate financial performance (CFP), R&D management, model misspecification*

1. Introduction

Analyzing the relationship between corporate social performance (CSP) and corporate financial performance (CFP) is part of a long-standing debate in management science (Cochran & Wood, 1984; Griffin & Mahon, 1997; Lu, Liu & Falkenberg, 2020; Margolis, Elfenbein, & Walsh, 2009; McGuire, Sundgren, & Schneeweis, 1988; Ullmann, 1985; Waddock & Graves, 1997). Essentially, the debate is driven by the ambiguity of the results. Over the years, researchers have identified positive, negative and neutral relationships between CSP and CFP (e.g., Delmas & Nairn-Birch, 2011; Feix & Philippe, 2020; Zhao & Murrell, 2016). Next to other factors, omitted variables, simultaneity, and measurement errors have been identified as one of the main drivers of this ambiguity (Hassan & Romilly, 2018).

McWilliams and Siegel (2000) addressed that omitted variables cause severe misspecifications in the underlying econometrical models and result in the mixed findings. By introducing the variables R&D intensity and advertising intensity, McWilliams and Siegel (2000) highlight how these important aspects had previously been omitted from research models. The inclusion of R&D intensity in their estimation model changed their results from a positive significant to an insignificant relationship between CSP and CFP. Leading to the conclusion that R&D intensity and CSP tend to mimic similar aspects of CFP.

However, after reviewing subsequent studies on this debate, we identified the emergence of two further empirical phenomena, which motivated us to conduct our study. First, following the econometrical re-specifications proposed by McWilliams and Siegel (2000), the ambiguity concerning the relationship between CSP-CFP should have eased in recent years. Yet we did not find more clarity in the results: despite integrating R&D intensity when investigating the CSP-CFP relationship, scholars still report positive (Lee, Cin, & Lee, 2016; Barnett & Salomon, 2012; Berrone, Surroca, & Tribó, 2007), neutral (Chang, Oh, & Messersmith, 2013; Darnall, Henriques, & Sadorsky, 2008), and negative results (Garcia-Castro, Ariño, & Canela, 2010;

Lioui & Sharma, 2012). Second, many studies have indeed found a significant relationship between R&D intensity and CFP – but pointing in a negative direction (Barnett & Salomon, 2012; Callan & Thomas, 2009; Erhemjamts, Li, & Venkateswaran, 2013). This second finding in particular conflicts with McWilliams and Siegel's (2000) underlying assumptions that increasing R&D intensity contributes to enhanced CFP.

Our article provides two main contributions to the existing literature. First, in order to disentangle the relationship between CSP and R&D intensity and its strategic implications, we emphasize econometrical issues arising from measurement errors. Analyzing different measurements for CSP, multiple studies have pointed out severe measurement errors resulting from limited construct validity and reliability (Orlitzky et al. 2003; Surroca and Tribo 2005; Mattingly and Berman 2006). While we can confirm that there is a neutral relationship between CSP and CFP for highly aggregated CSP measures, we find a significant positive relationship on a disaggregated level. In our case, CSP concerns are negatively associated with CFP. This finding supports arguments that scholars should refrain from using highly aggregated CSP scores as highly aggregated scores likely also cover similar aspects as other constructs, such as R&D intensity (Chen & Delmas, 2011; Rowley & Berman, 2000). Second, we highlight that the functional relationship between R&D intensity and CFP is not linear, but curvilinear. This finding highlights the importance of including R&D intensity as a non-linear variable (Bracker & Ramaya, 2011; Huang & Liu, 2005). On the one side, our contributions are important for researchers in order to acknowledge the impacts of measurement errors on the relationship between CSP and CFP as well as on the relation with other constructs. On the other side, we show that CSP and R&D intensity need to be managed separately as R&D intensity tends to mimic some aspects of CSP, but not all.

2. Method

As some of the first authors to highlight misspecification issues within the CSP-CFP debate, McWilliams and Siegel (2000) have been cited by more than 2,900 studies between 2000 and 2018. Using data from Compustat and KLD, they computed average annual values for all variables for the years 1991–1996, which resulted in a cross-sectional research model consisting of 524 firms. To measure CSP, McWilliams and Siegel (2000) used a binary variable (0;1) based on the affiliation with the index formerly known as the Domini 400 Social Index (DSI 400). This index was one of the first socially responsible investment indexes and is based on "ESG performance, sector alignment and size representation" (MSCI, 2015, p. 3). For CFP, McWilliams and Siegel only used accounting-based measures.

In alignment with the original study, we use the MSCI KLD database to compile CSP information for our study. MSCI KLD rates companies based on their social performance in 13 different categories. These evaluations are used to disclose a firm's social performance, as well as to construct social indices, such as the DSI 400. In order to avoid inconsistencies in the dataset, and because the rating methodology has changed over time, for our analysis we only use data from 2005–2013. We have obtained financial data from Compustat, resulting in an unbalanced panel of 927 firms with more than 3,500 firm year observations. In order to account for extreme values, we cut off all variables independently at the 5th and 95th percentiles to avoid bias from outliers (Flammer & Bansal, 2017; Zhao & Murrell, 2016).

Dependent variables

As an extension to McWilliams and Siegel (2000), who applied only accounting-based measures, we apply both accounting-based (Return on Assets, ROA) and market-based (Tobin's Q) measures of CFP. The accounting-based measures are based on ex-post accounting data. Therefore, this information can be regarded as backward looking. Market-based measures, on the other hand, are considered to capture a forward-looking view by integrating investor evaluations. ROA is calculated by dividing net income and total sales; Tobin's Q is based on the market value of equity plus the market value of debt over the book value of equity plus the book value of debt.

Independent variables

DSI 400

The first independent variable follows the design of McWilliams and Siegel's (2000) study and is based on their affiliation with the DSI 400, which is now known as the MSCI KLD 400 Social Index. Technically, this results in a dichotomous CSP variable: either a company is or is not part of the DSI 400. Since the McWilliams and Siegel study was published in 2000, only a limited number of further studies have considered the DSI 400 affiliation as a measure of CSP (see, for instance, Tsoutsoura, 2004).

KLD net score

Second, the KLD net score is calculated based on the difference between KLD strengths and concerns. Here, all strengths or concerns from the previously outlined assessment categories are aggregated to obtain a single score for each company. This aggregation results in a greater variation and has been frequently used by other scholars (Barnett & Salomon, 2012; Griffin & Mahon, 1997; Shi & Sun, 2015; Waddock & Graves, 1997). However, aggregating CSP data into a net score has also been widely criticized because performance differences are averaged out by this process (Chen & Delmas, 2011; Erhemjamts et al., 2013; Rowley & Berman, 2000). Additionally, there are empirical issues regarding the appropriateness of combining strengths and concerns without testing the validity of the factor aggregation (Mattingly & Berman, 2006). For our sample, we could obtain more than 3,500 observations with an average value of -0.3, ranging from -12 to 19.

KLD strengths and concerns

Third, in response to the aggregation issue, we account for more individuality in the CSP operationalization by considering KLD strengths and concerns separately (Burke, Hoitash, & Hoitash, 2019; Ioannou & Serafeim, 2015; Muller, 2018; Walls, Berrone, & Phan, 2012). For all KLD strengths, we obtain an average value of 1.7, ranging from 0 to 22, and for KLD concerns, we obtain an average value of 2.0, ranging from 0 to 18.

Control variables

We use controls for size, risk, R&D intensity, advertising intensity and industry affiliation based on the 4-digit SIC. Size is measured by total assets, and risk is measured by the ratio of debt to assets. R&D and advertising investments are divided by sales to compute the corresponding intensities. McWilliams and Siegel (2000) proposed that R&D intensity is a proxy for a firm's innovativeness. Here, innovativeness is characterized as an input variable that increases the technical knowledge capital and is not measured by potential outcomes (e.g., patents).

Research Model and robustness

For our study, we use the same set of variables as McWilliams and Siegel (2000) and extend their analysis by following a three-step approach. We start by replicating the McWilliams and Siegel (2000) model design by calculating average values for all variables over all years:

$$PERF_i = f(CSP_i, Size_i, Risk_i, IND_i, RDINT_i, INDADINT_i)$$

Our sample is based on a more recent database, from 2005–2013. This first step purely replicates the original McWilliams and Siegel (2000) analysis by using the affiliation with the DSI 400 to measure CSP.

For our second and third steps, we conduct an extended analysis by employing panel data regressions. Here, we apply the same reasoning, but use a firm- and time-fixed effects estimation to account for unobserved heterogeneity as well as for time effects. Following other recent studies, we also introduce further measurements of CSP to analyze differences that result from the variable design (Barnett & Salomon, 2012; Ioannou & Serafeim, 2015; Shi & Sun, 2015):

$$PERF_{i,t} = f(CSP_{i,t}, Size_{i,t}, Risk_{i,t}, R\&D_{i,t}, R\&D^2{}_{i,t}, INDADINT_{i,t} + Year_{i,t} + (\Omega Individual_i + \mu_{i,t})$$

Firms and time are denoted by i and t. In addition, we also include year dummies to control for time-fixed effects. $Individual_i$ represents the firm-fixed effect in order to control for time-invariant, firm-specific characteristics. The remaining error term captures all other influencing factors that are omitted through this setup. Beginning with pooled OLS, we tested for unobserved individual heterogeneity by applying the Breusch-Pagan Lagrange Multiplier test. Additionally, we performed a Hausman test to analyze whether the error terms are correlated with the other independent variables. The test results indicate that neither pooled OLS estimations nor random effects estimations are robust enough for this analysis. Hence, fixed effects estimation has been used for all models.

In the third step, we investigate how R&D intensity is related to CFP. Following McWilliams and Siegel (2000) and a common notion in the innovation literature (Hill & Snell, 1988; Rosenbusch, Brinckmann, & Bausch, 2011), the expected relationship between R&D intensity and CFP should be positive. However, many empirical studies have reported a negative relationship between R&D intensity and CFP (Barnett & Salomon, 2012; Callan & Thomas, 2009; Erhemjamts et al., 2013). Thus, we refer to more recent studies from innovation research and revise the functional form between R&D intensity and CFP, and test for a curvilinear relationship (Bracker & Ramaya, 2011; Huang & Liu, 2005).

3. Results

Table 1 presents a comparison of descriptive statistics between the original sample and the replication sample. For CFP, McWilliams and Siegel (2000) used an accounting-based variable. As there is no further specification, we employed ROA to compare basic summary statistics. In the original setup, CFP has a marginal negative mean value complemented by a comparably high standard deviation. In the replication sample, we found a moderate positive mean of 0.046 and a standard deviation of 0.051. The values for the DSI 400 affiliation remained quite stable over both samples. For R&D intensity, we identified a slightly higher mean value and a significantly lower standard deviation.

==

Insert Table 1 about here

==

Table 2 indicates the descriptive statistics for the extended sample in a panel structure. The values are on a similar level and range as those presented in table 1. In addition, we find that all correlations are statistically significant due to the large sample size.

==

Insert Table 2 about here

==

Table 3 reports the results for the replication and extended analyses. Model 1 displays the outcomes of the original study conducted by McWilliams and Siegel (2000). Model 2 indicates the results for the direct replication. For CSP, our findings adhere to the initial findings. The affiliation with the DSI 400 is not related to a firm's financial performance. R&D intensity, however, is moderately significant and points in a negative direction, contrary to the initial assumptions by McWilliams and Siegel (2000). Consequentially, this suggests that increasing R&D intensity is negatively associated with CFP. While this seems to be a surprising finding, other empirical studies report similar results (Barnett & Salomon, 2012; Callan & Thomas, 2009; Erhemjamts et al., 2013).

===

Insert Table 3 about here

===

In our extended analysis, we advance the original setup by integrating CSP variables with more variation and a curvilinear function for R&D intensity, and by controlling for unobserved heterogeneity in a fixed effects regression. For models 3 and 4, we use the KLD net score as our CSP variable, which can also be considered as a highly aggregated measurement. Here, we also find a neutral relationship for both market-based and accounting-based variables. Thus, this outcome further supports McWilliams and Siegel's hypothesis that CSP has a neutral effect on CFP when controlling for R&D intensity. Models 5 and 6 are based on more disaggregated CSP measures. Here, CSP strengths and concerns are separated. The results are different compared to those in the previous models. While the relationship between KLD strengths and both CFP variables remains neutral, we find a significant negative relationship for CSP concerns for ROA ($\rho < 0.01$) and for Tobin's Q ($p < 0.1$). This implies that a decrease in CSP concerns is positively related to CFP. Thus, analyzing more specific CSP constructs via disaggregated measures yields a better understanding of the CSP-CFP relationship. We can only find a neutral relationship for CSP variables that cover aggregated CSP aspects and CSP strengths, but not for CSP concerns.

The extended analyses (models 3–6) again report highly significant and negative coefficients for R&D intensity. However, all squared R&D intensities are also highly significant and point in a positive direction, i.e., following a u-shaped curve. Thus, we find support for a curvilinear relationship between R&D and CFP. Interestingly, the turning points of the functions for ROA, as well as for Tobin's Q, can be found at the upper end of the R&D intensity values. This implies that only very high values of R&D intensity are eventually related to improved financial performance.

===

Insert Table 4 about here

===

In order to test for potential influences of endogeneity in models 5 and 6, an instrumental variable estimation has been conducted. An appropriate instrument needs to be correlated with our CSP variable of interest and uncorrelated with the dependent variable. Accordingly, we use industry- and firm-related instruments in a two-stage least squares regression. Several scholars have used industry average CSP scores as their key instrumental variable (El Ghoul, Guedhami, Kwok, & Mishra, 2011; Koh, Qian, & Wang, 2014; Wang, Choi, & Li, 2008). Hence, we divide the firm-specific CSP score by the corresponding industry average CSP score in order to account for firm- and industry-specific systematic patterns (Garcia-Castro et al., 2010). Such an instrument can be considered as being "independent of any observable characteristics that affect the value of all firms in a given industry and year in the same manner" (Campa & Kedia, 2002, p. 1748). Additionally, we tested for the relevance and exogeneity of the outlined instrument. The under-identification test statistic (i.e., the Kleibergen-Paap test) yielded results with $p = 0.000$. Thus, the instrument appears to be sufficiently correlated with the endogenous variable. The weak identification statistics (i.e., Cragg-Donald Wald F-test) are greater than the 10 percent maximal IV size Stock-Yogo critical values, which confirms the relevance of our instrument (Stock & Yogo, 2005). Reviewing Hansen's J statistic ($p = 0.784$) to test for overidentifying restrictions, we also find support for our instrument selection, thus ensuring that our instrument is exogenous (Cameron & Trivedi, 2005). The instrument was then used in the second stage to re-estimate our research models. The results remained consistent over the instrumental variable estimation, as reported in table 4.

4. Discussion and Conclusion

This study revisits the CSP-CFP relationship and the role of R&D intensity in order to clarify sources of misspecification and improve strategical guidance to managers. McWilliams and Siegel (2000) prominently highlighted endogeneity problems in the underlying research models. Based on their stipulations, omitting R&D intensity from research models is expected to lead to an overestimation of CSP effects and wrong implications for managers; especially as CSP is a multidimensional latent construct that tends to cover similar aspects as R&D intensity. Our results confirm that it is important to control for R&D intensity when investigating the CSP-CFP relationship. However, we propose two important additional insights in order to prevent ambiguous results in empirical research settings and derive more focused strategies. First, our results corroborate the effects of measurement errors when measuring CSP. When analyzing the influence of R&D intensity, relying on one highly aggregated CSP measure results into serve concerns regarding construct validity. Although we find results similar to McWilliams and Siegel's (2000) when utilizing aggregated CSP measures (i.e., DSI 400; KLD net score), this is not the case under the revised model design when CSP is disaggregated into strengths and concerns as emphasized by other studies (Chatterji, Levine, & Toffel, 2009; Mattingly & Berman, 2006; Lee et al., 2016; Jayachandran, Kalaignanam, & Eilert, 2013; Kölbel, Busch, & Jancso, 2017). Consequentially, choosing the right level of aggregation is important to understand that CSP and R&D tend to cover similar aspects to a certain extent, but, ultimately, need to be addressed separately by corporate managers. One explanation for this finding is the fact that CSP is a broad latent construct that covers a wide range of areas (Erhemjamts et al., 2013). Various studies have used highly aggregated CSP measures to account for this multidimensionality (Chen & Delmas, 2011). Most prominently, affiliations with CSP indices and CSP scores have been used in this regard. However, the underlying measurement approach is highly disputed, and aggregating different aspects into one construct requires internal validity testing (Chatterji et al., 2009; Mattingly & Berman, 2006). This is particularly true in extreme cases (e.g., dichotomization into binary variables) where the loss of variation suppresses an assessment of the multidimensional character of CSP, and effect sizes are likely to be underestimated for the CSP-CFP relationship (Burke et al., 2019; Dawson & Weiss, 2012; MacCallum, Zhang, Preacher, & Rucker, 2002). Accordingly, disaggregated CSP variables have been used increasingly in recent studies (Hubbard, Christensen, & Graffin, 2017; Ioannou & Serafeim, 2015). CSP strengths are based on an evaluation of management strategies regarding a company's "innovation capacity […] and risk management" (MSCI, 2016, pp. 19), thus based on compliance with basic sustainability standards and systems. Measurement of CSP concerns is mainly based on an assessment of "ESG controversies" (MSCI, 2016, p. 14). Our results indicate that there is a statistically significant relationship between CSP concerns and CFP, but not between aggregated CSP measures as well as CSP strengths and CFP. One interpretation of this outcome is that CSP concerns can be regarded as visible CSP outcomes based on how well process and systems have been implemented (Ameer & Othman, 2012). Thus, on the one hand, we emphasize that R&D and CSP concerns need to be handled separately by companies as both factors significantly influence the financial performance. On the other hand, aggregated CSP measures need to be handled carefully as they tend to mimic similar strategic directions like R&D. Accordingly, the right level of aggregation needs to be carefully analyzed to derive meaningful strategic decisions.

Second, we stress that the functional form of R&D intensity and CFP is more complex than hypothesized (McWilliams & Siegel, 2000). Assuming a linear function for R&D intensity tends to be a source of further misspecifications. This can be illustrated by multiple studies that have reported a significant negative relationship between R&D intensity and CFP (Barnett & Salomon, 2012; Callan & Thomas, 2009; Erhemjamts et al., 2013). In other words, this outcome would lead to the suggestion to minimize R&D intensity. Our results show that the relationship between R&D and CFP is actually more complex and dependent on multiple aspects (Cabral, 2003). Corresponding with Bracker and Ramaya (2011) and Huang and Liu (2005), we identify a curvilinear relationship between R&D intensity and CFP. In contrast to these previous studies, we find the curvilinear relationship to be u-shaped. The difference in the results is related to the treatment of R&D intensity. Previous studies often used industry related R&D intensity as suggested by the industrial organization (IO) literature and due to limited data availability (Bracker & Ramaya, 2011; Huang & Liu, 2005). Following these assumptions, there would be no firm-specific heterogeneity between companies concerning their R&D intensity, and R&D intensity would only differ between industries. However, recent studies that follow the resource-based view analyze R&D intensity on an individual firm level, and acknowledge that the differences between firms lead to superior financial performance (Bromiley, Rau, & Zhang, 2017; Sujit & Mukherjee, 2005).

Bibliography

Ameer, R., & Othman, R. (2012). Sustainability Practices and Corporate Financial Performance: A Study Based on the Top Global Corporations. *Journal of Business Ethics*, 108(1), 61–79. https://doi.org/10.1007/s10551-011-1063-y

Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33(11), 1304–1320. https://doi.org/10.1002/smj.1980

Berrone, P., Surroca, J., & Tribó, J. A. (2007). Corporate Ethical Identity as a Determinant of Firm Performance. A Test of the Mediating Role of Stakeholder Satisfaction. *Journal of Business Ethics*, 76(1), 35–53. https://doi.org/10.1007/s10551-006-9276-1

Bracker, K., & Ramaya, K. (2011). Examining the impact of research and development expenditures on Tobin's Q. *Academy of Strategic Management Journal*, 10, 63-79.

Bromiley, P., Rau, D., & Zhang, Y. (2017). Is R & D risky? *Strategic Management Journal*, 38(4), 876-891. https://doi.org/10.1002/smj.2520

Burke, J. J., Hoitash, R., & Hoitash, U. (2019). The Heterogeneity of Board-Level Sustainability Committees and Corporate Social Performance. *Journal of Business Ethics*, 154(4), 1161-1186. https://doi.org/10.1007/s10551-017-3453-2

Cabral, L. M. (2003). R&D competition when firms choose variance. *Journal of Economics & Management Strategy*, 12(1), 139-150. https://doi.org/10.1111/j.1430-9134.2003.00139.x

Callan, S. J., & Thomas, J. M. (2009). Corporate financial performance and corporate social performance. An update and reinvestigation. *Corporate Social Responsibility and Environmental Management*, 16(2), 61–78. https://doi.org/10.1002/csr.182

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge, U.K.: Cambridge University Press.

Campa, J. M., & Kedia, S. (2002). Explaining the diversification discount. *The Journal of Finance*, 57(4), 1731–1762. https://doi.org/10.1111/1540-6261.00476

Chang, Y. K., Oh, W. Y., & Messersmith, J. G. (2013). Translating corporate social performance into financial performance: exploring the moderating role of high-performance work practices. *International Journal of Human Resource Management*, 24(19), 3738–3756. https://doi.org/10.1080/09585192.2013.778312

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, 18(1), 125-169. https://doi.org/10.1111/j.1530-9134.2009.00210.x

Chen, C.-M., & Delmas, M. (2011). Measuring Corporate Social Performance. An Efficiency Perspective. *Production and Operations Management,* 20(6), 789–804. https://doi.org/10.1111/j.1937-5956.2010.01202.x

Cochran, P. L., & Wood, R. A. (1984). Corporate social responsibility and financial performance. *Academy of Management Journal,* 27(1), 42–56. https://doi.org/10.5465/255956

Darnall, N., Henriques, I., & Sadorsky, P. (2008). Do environmental management systems improve business performance in an international setting? *Journal of International Management,* 14(4), 364–376. https://doi.org/10.1016/j.intman.2007.09.006

Dawson, N. V., & Weiss, R. (2012). Dichotomizing continuous variables in statistical analysis: A practice to avoid. *Medical Decision Making,* 32(2), 225-226. https://doi.org/10.1177/0272989X12437605

Delmas, M. A., & Nairn-Birch, N. S. (2011). Is the tail wagging the dog? An empirical analysis of corporate carbon footprints and financial performance. *Institute of the Environment and Sustainability*.

El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance,* 35(9), 2388–2406. https://doi.org/10.1016/j.jbankfin.2011.02.007

Erhemjamts, O., Li, Q., & Venkateswaran, A. (2013). Corporate Social Responsibility and Its Impact on Firms' Investment Policy, Organizational Structure, and Performance. *Journal of Business Ethics,* 118(2), 395–412. https://doi.org/10.1007/s10551-012-1594-x

Feix, A., & Philippe, D. (2020). Unpacking the Narrative Decontestation of CSR: Aspiration for Change or Defense of the Status Quo? *Business & Society,* 59(1), 129–174. https://doi.org/10.1177/0007650318816434

Flammer, C., & Bansal, P. (2017). Does a long-term orientation create value? Evidence from a regression discontinueity. *Strategic Management Journal,* 38(9), 1827-1847. https://doi.org/10.1002/smj.2629

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), 107–126. https://doi.org/10.1007/s10551-009-0143-8

Griffin, J. J., & Mahon, J. F. (1997). The Corporate Social Performance and Corporate Financial Performance Debate. *Business & Society,* 36(1), 5–31. https://doi.org/10.1177/000765039703600102

Hassan, O. A., & Romilly, P. (2018). Relations between corporate economic performance, environmental disclosure and greenhouse gas emissions: New insights. *Business Strategy and the Environment,* 27(7), 893-909. https://doi.org/10.1002/bse.2040

Hill, C. W. L., & Snell, S. A. (1988). External control, corporate strategy, and firm performance in research-intensive industries. *Strategic Management Journal,* 9(6), 577–590. https://doi.org/10.1002/smj.4250090605

Huang, C., & Liu, C. (2005). Exploration for the relationship between innovation, IT and performance. *Journal of Intellectual Capital,* 6(2), 237–252. https://doi.org/10.1108/14691930510592825

Hubbard, T. D., Christensen, D. M., & Graffin, S. D. (2017). Higher highs and lower lows: The role of corporate social responsibility in CEO dismissal. *Strategic Management Journal,* 38(11), 2255-2265. https://doi.org/10.1002/smj.2646

Ioannou, I., & Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. *Strategic Management Journal*, 36(7), 1053-1081. https://doi.org/10.1002/smj.2268

Jayachandran, S., Kalaignanam, K., & Eilert, M. (2013). Product and environmental social performance: Varying effect on firm performance. *Strategic Management Journal, 34*(10), 1255-1264. https://doi.org/10.1002/smj.2054

Kölbel, J. F., Busch, T., & Jancso, L. M. (2017). How media coverage of corporate social irresponsibility increases financial risk. *Strategic Management Journal,* 38(11), 2266-2284. https://doi.org/10.1002/smj.2647

Koh, P.-S., Qian, C., & Wang, H. (2014). Firm litigation risk and the insurance value of corporate social performance. *Strategic Management Journal,* 35(10), 1464–1482. https://doi.org/10.1002/smj.2171

Lee, K. H., Cin, B. C., & Lee, E. Y. (2016). Environmental responsibility and firm performance: the application of an environmental, social and governance model. *Business Strategy and the Environment,* 25(1), 40-53. https://doi.org/10.1002/bse.1855

Lioui, A., & Sharma, Z. (2012). Environmental corporate social responsibility and financial performance. Disentangling direct and indirect effects. *Ecological Economics,* 78, 100–111. https://doi.org/10.1016/j.ecolecon.2012.04.004

Lu, H., Liu, X., & Falkenberg, L. (2020). Investigating the Impact of Corporate Social Responsibility (CSR) on Risk Management Practices. *Business & Society.* https://doi.org/10.1177/0007650320928981

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods,* 7(1), 19-40. https://doi.org/10.1037/1082-989X.7.1.19

Margolis, J. D., Elfenbein, H. A., & Walsh, J. P. (2009). Does it pay to be good… and does it matter? A meta-analysis of the relationship between corporate social and financial performance. *SSRN Electronic Journal.* http://dx.doi.org/10.2139/ssrn.1866371

Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, 45(1), 20-46. https://doi.org/10.1177/0007650305281939

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of management Journal,* 31(4), 854–872. https://doi.org/10.5465/256342

McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal,* 21(5), 603–609. https://doi.org/10.1002/(SICI)1097-0266(200005)21:5<603::AID-SMJ101>3.0.CO;2-3

MSCI (eds.). (2015). MSCI KLD 400 Social Index fact sheet. https://www.msci.com/documents/10199/904492e6-527e-4d64-9904-c710bf1533c6

MSCI (eds.). (2016). MSCI ESG KLD STATS: 1991-2015 DATA SETS. Methodology. https://www.wiso.uni-hamburg.de/bibliothek/recherche/datenbanken/unternehmensdaten/msci-methodology-2014.pdf

Muller, A. (2018). When Does Corporate Social Performance Pay for International Firms? *Business & Society.* https://doi.org/10.1177/0007650318816957

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies,* 24(3), 403-441. https://doi.org/10.1177/0170840603024003910

Rosenbusch, N., Brinckmann, J., & Bausch, A. (2011). Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing,* 26(4), 441–457. https://doi.org/10.1016/j.jbusvent.2009.12.002

Rowley, T., & Berman, S. (2000). A brand new brand of corporate social performance. *Business & Society,* 39(4), 397–418. https://doi.org/10.1177/000765030003900404

Shi, G., & Sun, J. (2015). Corporate Bond Covenants and Social Responsibility Investment. *Journal of Business Ethics,* 131(2), 285–303. https://doi.org/10.1007/s10551-014-2272-y

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge and New York: Cambridge University Press, 80–108. https://doi.org/10.1017/CBO9780511614491.006

Surroca, J., & Tribó, J. A. (2005). The Corporate Social and Financial Performance Relationship: What's the Ultimate Determinant?. *Academy of Management Proceedings* (Vol. 2005, No. 1, pp. E1-E6). https://doi.org/10.5465/ambpp.2005.18783433

Sujit, K. S., & Mukherjee, B. (2005). *How Much Does R&D decision depend on firm, industry, group and its interactions* (No. 0503008). University Library of Munich, Germany.

Tsoutsoura, M. (2004). Corporate Social Responsibility and Financial Performance, Center for Responsible Business. Working Paper Series, Paper 7, *University of California*. Available at https://escholarship.org/uc/item/111799p2

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 US firms. *Academy of management review,* 10(3), 540–557. https://doi.org/10.5465/amr.1985.4278989

Waddock, S. A., & Graves, S. B. (1997). The corporate social performance-financial performance link. *Strategic Management Journal*, 18(4), 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G

Walls, J. L., Berrone, P., & Phan, P. H. (2012). Corporate governance and environmental performance: Is there really a link? *Strategic Management Journal,* 33(8), 885-913. https://doi.org/10.1002/smj.1952

Wang, H., Choi, J., & Li, J. (2008). Too little or too much? Untangling the relationship between corporate philanthropy and firm financial performance. *Organization Science*, 19(1), 143–159. https://doi.org/10.1287/orsc.1070.0271

Zhao, X., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, 37(11), 2378–2388. https://doi.org/10.1002/smj.2579

Tables

Table 1: *Descriptive statistics for the original and replication model*

	Original sample (McWilliams & Siegel, 2000)					Replication sample				
Variable	Mean	Std Dev	CFP	DSI400	RDINT	Mean	Std Dev	CFP	DSI400	RDINT
Accounting-based CFP	-0.011	1.043	1.000			0.046	0.051	1.000		
DSI400	0.619	0.345	0.356**	1.000		0.681	0.320	0.146***	1.000	
R&D intensity	0.011	0.949	0.403**	0.449***	1.000	0.027	0.053	0.029	0.1047**	1.000

All variables computed as annual averages over the respective period.
*p ≤ 0.1; **p ≤ 0.05; ***p ≤ 0.01

Table 2: *Descriptive statistics for the extended models*

	Replication sample							
Variable	Mean	Std Dev	ROA	Tobin's Q	KLD net score	KLD strengths	KLD concerns	RDint
ROA	0.029	0.065	1.000					
Tobin's Q	0.996	0.881	0.332***	1.000				
KLD net score	-0.291	2.673	0.072***	0.040***	1.000			
KLD strengths	1.553	2.543	0.125***	0.023***	0.751***	1.000		
KLD concerns	1.845	1.846	0.069***	-0.024***	-0.414***	0.291***	1.000	
R&D intensity	0.030	0.061	-0.051***	0.036***	0.090***	0.054***	-0.051***	1.000

*p ≤ 0.1; **p ≤ 0.05; ***p ≤ 0.01

Table 3: *Revisiting the role of R&D intensity on CSP-CFP*

VARIABLES	(1) McWilliams &Siegel	(2) Replication	(3) ROA	(4) TQ	(5) ROA	(6) TQ
DSI400	-0.062	0.0106				
	(0.059)	(0.0126)				
Score			0.000499	0.00804		
			(0.000588)	(0.00497)		
Strengths					-0.000725	0.00273
					(0.000708)	(0.00668)
Concerns					-0.00272***	-0.0178*
					(0.00104)	(0.00953)
Size	Included	0.0081**	0.00439	-0.404***	0.00575	-0.399***
		(0.0036)	(0.00518)	(0.0564)	(0.00522)	(0.0558)
Risk	Included	-0.0035	-0.00912***	-0.110***	-0.00915***	-0.110***
		(0.0055)	(0.00147)	(0.0149)	(0.00148)	(0.0149)
RDint	0.263***	-0.2676*	-0.720***	-4.019***	-0.719***	-4.005***
	(0.050)	(0.1478)	(0.0800)	(1.147)	(0.0800)	(1.142)
$Rdint^2$			0.262***	4.573**	0.261***	4.559**
			(0.0575)	(1.913)	(0.0576)	(1.905)
ADint	Included	0.1340	-0.0997	-2.995	-0.0666	-2.874
		(0.4002)	(0.401)	(4.585)	(0.400)	(4.566)
Constant	-	-0.1150***	0.0481	4.393***	0.0448	4.385***
		(0.0301)	(0.0371)	(0.408)	(0.0371)	(0.406)
Observations	524	452	3,839	3,506	3,839	3,506
R-squared	0.29	0.579	0,195	0.244	0,197	0.245
No. of firms	524	452	927	851	927	851
Industry Dummies	YES	YES	NO	NO	NO	NO
Firm FE	NO	NO	YES	YES	YES	YES
Time FE	NO	NO	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: *Results of the Instrumental Variable Estimation*

VARIABLES	(7) ROA	(8) TQ
IV Concerns	-0.002**	-0.033***
	(0.001)	(0.008)
Size	0.007**	-0.341***
	(0.004)	(0.050)
Risk	-0.009***	-0.120***
	(0.001)	(0.014)
RDint	-0.749***	-0.687
	(0.072)	(0.548)
$Rdint^2$	0.280***	0.525
	(0.048)	(0.423)
ADint	-0.154	-6.197**
	(0.331)	(2.621)
Observations	3,640	3,732
R-squared	0,149	0,086
No. of firms	728	740
Firm FE	YES	YES
Time FE	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Annex 3: Third Research Paper

A change will do you good:
Do continual environmental improvements matter?

Research Abstract: *Numerous studies have utilized corporate environmental performance (CEP) to investigate its association with corporate financial performance (CFP). Many studies use aggregated environmental scores to represent CEP. This article examines whether continual CEP improvements of operational environmental indicators are reflected in environmental scores provided by sustainability rating agencies. Furthermore, we investigate whether continual CEP improvements affect CFP. Based on panel data (2008–2016) with more than 3,000 firms, the results show that continual CEP improvements are not associated with aggregated environmental scores. Interestingly, these improvements are positively associated with accounting- and market-based CFP. The article concludes that continual CEP improvements are relevant from a materiality perspective, but not adequately captured in sustainability ratings. Thus, continual CEP improvements should be integrated more prominently in research and practice.*

Managerial Abstract: *Corporate managers, financial investors, and academic scholars have long track records of considering corporate environmental performance (CEP) as an indication of how companies handle external opportunities and risks dealing with the natural environment. While previous research has considered various organizational capabilities related to environmental responsiveness, one important capability has remained under-scrutinized: continual CEP improvements. By focusing on improvements of operational environmental indicators, our results find that according improvements are not adequately reflected in aggregated environmental scores provided by sustainability rating agencies. This is a severe drawback since CEP improvements are highly relevant from a materiality perspective. Based on our results, we encourage managers and investors to actively promote and utilize continual CEP improvements as a strategic capability in company assessments.*

Keywords: corporate environmental performance, continual improvement, environmental scores, financial performance, organizational capabilities

1. Introduction

Over 40 years, scholars have utilized corporate environmental performance (CEP) in empirical research settings, e.g., to investigate its association with corporate financial performance (CFP) (Grewatsch & Kleindienst, 2017; Hang, Geyer-Klingeberg, & Rathgeber, 2018). Previous studies describe CEP as companies' cumulative policies, processes, and activities to manage and reduce environmental impacts (Trumpp, Endrikat, Zopf, & Guenther, 2015). This research has dealt broadly with definitions (*what*) and motivations (*why*) of CEP, especially in its association with CFP (Zollo, Cennamo, & Neumann, 2013). Recently, more attention has been dedicated to the temporal dimension (*when*) in the CEP-CFP relationship. While many studies propose that superior CEP does not hurt a company's financial bottom line (Eccles, Ioannou, & Serafeim, 2014; Endrikat, Guenther, & Hoppe, 2014; Flammer, 2015), recent findings propose that superior levels of CEP have long-term payoffs rather than short-term or intermediate benefits (Alvarez, 2012; Delmas, Nairn-Birch, & Lim, 2015; Hang et al., 2018; Misani & Pogutz, 2015).

Research dealing with strategic aspects of ecological responsiveness has suggested that a proactive environmental strategy can positively affect CEP and CFP simultaneously, which offers promising correlation results (Russo & Fouts, 1997). Previous works on proactive environmental strategies found that companies with proactive environmental strategies possess certain intangible resources and strategic capabilities (Aragón-Correa & Sharma, 2003; Hart, 1995; Russo & Fouts, 1997; Sharma & Vredenburg, 1998; Surroca, Tribó, & Waddock, 2010; Walls, Phan, & Berrone, 2011). This literature highlights several key capabilities, inter alia, shared vision, employee involvement, stakeholder integration, and continuous improvements, which help companies realize economic benefits from environmentally sound practices. While previous studies have incorporated many organizational capabilities in the CEP-CFP context (Russo & Fouts, 1997; Sharma & Vredenburg, 1998; Surroca et al., 2010), these studies do not portray CEP as the rate of continual improvement. Rather, these studies often measure CEP as an aggregated level of CEP resembling corporate ecological responsiveness (Bansal & Roth, 2000). In this paper, we focus on continual CEP improvement – referring to the definition from the International Organization for Standardization (ISO, 2015) – as companies' "recurring activity to enhance environmental performance".

Several longitudinal analyses have considered change-related properties of CEP, e.g., by focusing on individual elements of CEP, such as carbon emissions (Alvarez, 2012; Delmas et al., 2015; Lewandowski, 2017) or toxic waste (Berchicci, Dowell, & King, 2012; 2017). The findings provide novel insights into spillover effects from mergers and acquisitions (Berchicci et al., 2017), conditions of competitiveness and profitability under regulatory uncertainty (Delmas et al., 2015), and firms' levels of resiliency, or the ability to cope with unexpected situations over longer periods (Ortiz-de-Mandojana & Bansal, 2016). Yet, no study has investigated whether continual CEP improvements at the operational level are adequately captured by aggregated environmental scores provided by sustainability rating agencies, or how such CEP improvements are related to different measures of financial performance. Even panel data analyses that include CEP data across several years have not really offered a dynamic view of CEP. Based on multiple samples covering the years 2008–2016, we seek to answer two research questions: (1) How are continual CEP improvements reflected in aggregated environmental scores provided by sustainability rating agencies?; and, (2) To what degree do continual CEP improvements affect CFP?

To operationalize continual CEP improvements, we gathered data on environmental performance indicators at the operational level, including environmental inputs (i.e., energy use and water use) and environmental outputs (i.e., carbon emissions and water discharge). The results show that continual CEP improvements are not associated with aggregated CEP scores in a meaningfully way. This holds true for the aggregated KLD environment score as well as KLD scores covering environmental strengths and concerns separately. These findings are confirmed in a robustness check using a completely different environmental score, the Asset4 Environmental Score. Nevertheless, our results show that all measures of continual CEP improvement are positively associated with both measures of CFP, including return on assets (ROA) and Tobin's q.

This study contributes to the strategic management literature in two respects. First, recent studies have investigated the validity and consistency of environmental data provided by sustainability rating agencies (Berg, Koelbel, & Rigobon, 2019; Chatterji, Durand, Levine, & Touboul, 2016; Delmas, Etzion, & Nairn-Birch, 2013; Eccles, Lee, & Strohle, 2019). We are able to contribute to this literature by demonstrating that continual CEP improvements are not adequately reflected in environmental scores, including KLD and Asset4. This finding highlights that continually improving CEP on the operational level do not automatically factor into the rating agencies' environmental scores. This would accentuate the inability of environmental scores to capture dynamic environmental progress (Delmas et al., 2013). Second, previous literature has argued whether continual CEP improvements should be considered a valuable capability (Aragon-Correa & Sharma, 2003; Delmas & Toffel, 2008; Sharma & Vredenburg, 1998; Walls et al., 2011). Operational environmental indicators, including energy use, water use, carbon emissions, and water discharge, lend themselves to the improvement measures very fittingly. Thus, managers, investors, and academics should not rely purely on aggregated environmental scores provided by sustainability rating agencies, but should integrate continual improvement indices to create a more holistic view of companies' CEP.

The article is structured as follows: the next section summarizes the exact literature on CEP, including the various dimensions of CEP. Afterwards, we derive hypotheses regarding the relevance of continual CEP improvements. The methods section explains our dataset and variables. The findings provide the results in accordance with the suggested hypotheses and robustness checks. Lastly, we discuss these results within the context of strategic management literature and conclude with several implications for future research and strategic management.

2. Theoretical Background and Related Literature

2.1. Multidimensionality of CEP

CEP is a multidimensional construct with distinct, albeit interrelated, aspects. According to Ilinitch, Soderstrom, and Thomas (1998, p. 386), "defining corporate environmental performance is not a straightforward task", which has been plagued with conceptual ambiguity (Suddaby, 2010) and incompatible results between studies due to the use of inconsistent measures (Etzion, 2007). Based on a systematic literature review of CEP-related articles, Trumpp et al. (2015) offered an abridged understanding of CEP according to the ISO 14031, a norm for environmental performance evaluation. According to the ISO norm (2013), CEP is defined as "measurable results of an organization's management of its environmental aspects". While this definition delivers conceptual clarity, the operationalization of CEP remains under the discretion of the investigator (i.e., academic researcher, investment analyst, etc.), who applies a designated measure or proxy of CEP.

For this reason, several authors have attempted to classify various elements, or dimensions, of CEP. Ilinitch et al. (1998) proposed four categories of CEP based on two dimensions (process and outcome), including internal systems and stakeholder relations as process-related categories, and internal (regulatory) compliance and external (environmental) impacts as outcome-related categories. Trumpp et al. (2015) streamlined these suggestions into two main categories, environmental management performance (EMP) and environmental operational performance (EOP), which are consistent with the ISO 14031 norm. As a process-related category, EMP contains the "strategic level of environmental performance and focuses on management principles and activities with regard to the natural environment". As an outcome-related category, EOP includes "the environmental impacts of a firm's management activities regarding the natural environment" (Trumpp et al., 2015, p. 190).

A common method for investigating the association of CEP to financial performance measures is the consideration of cross-sectional or pooled data for a single given year. This can be considered the static view of CEP (Delmas et al., 2015). For example, static-based CEP studies would investigate if firms with higher environmental scores experience greater profitability in the same year as compared to firms with lower environmental scores (Certo, Withers, & Semadeni, 2017). However, this static-based view has several drawbacks. First, static studies on the CEP-CFP relationship use limited datasets, which are susceptible to time-invariant unobservable heterogeneity (Surroca et al., 2010). For example, Elsayed and Paton (2005, p. 398) warn, "it is well known that inferences based on such cross-sectional analyses are likely to be invalid in the presence of significant firm heterogeneity." Second, static data can only explain correlation; it does not provide any indication of the direction of causality (Surroca et al., 2010; Hang et al., 2019). Third, the static view compares levels of CEP between firms, but cannot deduce if improvements to CEP will have any effect on the results (Elsayed & Paton, 2005; Slawinski & Bansal, 2015). According to Delmas et al. (2013, p. 376), the static-based view of CEP "has offered little to explain the dynamic interactions between proactive environmental strategies and financial performance."

2.2. Temporal dimensions of CEP

In response to the concerns associated with the static-based approach in former CEP studies, including firm heterogeneity and missing indications of causality, temporal dimensions have been incorporated into CEP studies according to various methods. First, cross-sectional studies have increasingly embraced time lags, most commonly lagging the dependent variable, i.e., CFP (Hart & Ahuja, 1996; Russo & Fouts, 1997; King & Lenox, 2002). By lagging CFP, studies have revealed that CEP can have a positive influence on CFP (Dixon-Fowler, Slater, Johnson, Ellstrand, & Romi, 2013). Temporal studies are able to differentiate the effect that CEP imposes on several CFP indicators, including accounting-based and market-based performance with multiple time lags (Cheng, Ioannou, & Serafeim, 2014; Horváthová, 2012; Lundgren & Zhou, 2017; Trumpp & Guenther, 2017). However, Hang et al. (2019, p. 257) emphasized that although "delayed effects are not identical to causality, they are at least accepted as a strong indicator."

Second, longitudinal studies have flourished in recent years to investigate the CEP-CFP relationship using common variables for multiple years. Typically, these studies have included CEP data provided by sustainability rating agencies, i.e., aggregated environmental scores per year are used as the independent variable and regressed on measures of CFP (Alvarez, 2012; Wagner, 2010; Yadav, Han, & Kim, 2017). Studies in this area do not necessarily reflect improvements, as they just relate the level of CEP of a given year to the level of CFP for the same year or a lagged year. The advantage herein is specified by Elsayed and Paton (2005, p. 298), as "panel data allows researchers to control for unobservable firm-specific effects and, as a consequence, has the potential to provide a much more powerful evidence base."

Third, longitudinal studies have manipulated CEP data in a dynamic way, i.e., by showing the change in environmental variables, mostly focusing on environmental scores. While the dynamic nature of CEP has been mentioned in earlier studies (e.g., Russo & Fouts, 1997), it has only recently been measured and empirically tested (Berchicci et al., 2012; 2017; Delmas et al., 2015; Short et al., 2016). However, there is no consistent approach to creating dynamic CEP variables, as existing studies operationalize CEP differently: as a *change* in facilities' waste generation (Berchicci et al., 2017), as a *decrease* in firms' greenhouse gas emissions (Delmas et al., 2015), or as an *improvement* of firms' environmental scores (Alvarez, 2012; Yadav et al., 2017). Additionally, each study focuses on a single, albeit dissimilar environmental variable. Previous studies have not investigated whether continual CEP improvements are adequately captured by aggregated environmental scores provided by sustainability rating agencies, or how multiple variables reflecting continual CEP improvements on the operational level are related to different measures of financial performance.

2.3. Continual improvement as an organizational capability

The notion of continual improvement originates in the quality management literature (Deming, 1986), and it has proliferated through lean production and Six Sigma initiatives (Anand, Ward, Tatikonda, & Schilling, 2009; Voss, 2005). Continual improvement describes firms' ability to constantly modify their planning, implementation, and controlling mechanisms with the goal of enhancing performance. As many organizational fields become increasingly complex and experience rapid technological changes, firms do not compete on superior organizational structure alone; they also rely on their capability to improve the quality of their existing products, services, and processes on a continual basis (Anand et al., 2009; Teece, 2007). As the concept of continual improvement implies that firms constantly adjust, it contests prior theories of organizational change that view incessant modifications "as threats to the system, which attempts to return to a state of dynamic equilibrium" (Ledford, 2015, p. 1).

In an ecological context, the ISO 14001 – an internationally renowned norm for environmental management – promotes continual improvement as the main, substantive construct to reduce environmental impacts within the management system (Brouwer & van Koppen, 2008; Delmas & Toffel, 2008). On a strategic level, the pursuit of continual improvement enables managers to plan, implement, execute, and control specifically defined environmental strategies on a long-term basis (Hart, 1995; Helfat & Winter, 2011). On an operational level, continual improvements encourage action plans specifically aimed at increasing reductions of environmental outcomes on an annual basis (Delmas & Toffel, 2008). According to international environmental standards, especially the ISO 14001, key operational indicators to assess continual improvement include energy use, water use, carbon emissions, water discharge, and waste (Brouwer & van Koopen, 2008).

Building upon theoretical perspectives of proactive environmental strategies (Bansal & Roth, 2000; Buysse & Verbeke, 2003) and corresponding organizational capabilities (Aragon-Correa & Sharma, 2003; Hart, 1995; Russo & Fouts, 1997; Sharma & Vredenburg, 1998), the competitive advantage of environmentally proactive firms is likely attributable to several complementary assets and capabilities (Christmann, 2000; Hang et al., 2019; Hart, 1995). An organizational capability is described as a company's ability to perform a coordinated set of tasks using its given resources to achieve a desired result (Helfat & Peteraf, 2003). Continual improvement is one of these capabilities, which can lead to the achievement of multiple results simultaneously. On the one hand, it can boost competitive advantages attained through economic value-added activities and outcomes, including better utilization of inputs and more efficient processes. On the other hand, it can accelerate cost savings via the reduction of natural resources and wastes. Research has shown that companies possessing the capability for continual improvement, from existing quality management and pollution prevention systems, have lower implementation costs for environmental management systems (Darnall & Edwards, 2006; Hart & Milstein, 2003). In turn, these positive trends can lead to further spillover effects in other areas, including increased investor confidence and a positive reputation with internal and external stakeholders (Hart & Dowell, 2011).

3. Hypotheses Development

3.1. Continual CEP improvements and aggregated environmental scores

Due to the rising demand for socially and environmentally responsible investment products over the last years, a significant interest in reliable environmental data has emerged (Delmas et al., 2013; Eccles et al., 2019). This has led to the creation and expansion of proprietary data providers and sustainability rating agencies offering a range of data-related services. Well-known players in this field include the Dow Jones Sustainability Index (López, Garcia, & Rodriguez, 2007), the Domini 400 Social Index (McWilliams & Siegel, 2000; Simpson & Kohers, 2002), the Carbon Disclosure Project (Delmas et al., 2015; Misani & Pogutz, 2015), and the Kinder, Lydenberg, and Domini (KLD) score (Chatterji & Levine, 2006; Chen and Delmas, 2011; Eccles et al., 2019; Mattingly & Berman, 2006; Surroca et al., 2010).

Recent studies have demonstrated that the underlying methodologies for generating aggregated environmental scores diverge to the point that the results of different sustainability rating agencies can hardly be compared (Berg et al., 2019; Chatterji et al., 2016; Eccles et al., 2019). The KLD environmental score is regarded as exceptional in its ability to explain a good portion of variance between firms' environmental management policies on the one hand, and their environmental outcomes and impacts on the other hand (i.e., it has been found to explain 80% of variance between these variables; Delmas et al., 2013; Misani & Pogutz, 2015).

As one of the most commonly used measures of CEP in academic research, KLD provides an amassed sum of various performance indicators, including assessments of environmental strategies, policies, processes, and outcomes. KLD Research & Analytics Inc., which has been owned by MSCI ESG since 2010, collects data via an impartial and objective rating methodology, which is considered a valid, reliable construct for CEP (Surroca et al., 2010). KLD separates environmental indicators according to strengths and concerns, which are considered separate but interrelated constructs (Mattingly & Berman, 2006; Semenova &

Hassel, 2015). According to KLD (2015), positive performance indicators (i.e., strengths) are rated as "1" for meeting the assessment criteria, "0" for not meeting the criteria, and "NR" for not covered, whereas negative performance indicators (i.e., concerns) are rated as "1" for raising substantial external concerns, "0" for raising no concerns, and "NR" for not researched.

Previous studies have used KLD scores in two main ways. A first approach aggregates all environmental indicators, and calculates an environmental net score by subtracting the environmental concerns from environmental strengths (Barnett & Salomon, 2012; Mattingly, 2017). The KLD environmental score provides greater coverage of multiple aspects, but it raises validity issues due to combining various unrelated and non-convergent constructs (Mattingly & Berman, 2006; Perrault & Quinn, 2018; Tsai & Wang, 2005). A second approach separates the two dimensions KLD environmental strengths and KLD environmental concerns (Berman, 2006; Delmas et al., 2015; Erhemjamts, Li, & Venkateswaran, 2013; Perrault & Quinn, 2018). Previous studies have also suggested conveying all scores separately, including net score, strengths, and concerns. Semenova and Hassal (2015, p. 251) stress that "KLD environmental strengths and concerns [indeed] converge on the same latent factor of institutional weakness and are significantly positively correlated". Therefore, KLD is considered a suitable measure that should be able to capture continual CEP improvements.

Continual CEP improvements in a firm's operations – such as increased resource productivity and decreased carbon emissions – aim to reduce the environmental impact (Schultze & Trommer, 2012) and have a positive effect on a firm's overall environmental performance. Thus, we assume that continual CEP improvements will enhance environmental performance scores. Furthermore, CEP improvements can reduce environmental related risks and, thus, should lead to increased reputation and investor confidence in these firms (Hart & Dowell, 2011; Semenova & Hassel, 2015). Overall, continual CEP improvements on the operational level should have an overall positive association with environmental scores, which are provided by sustainability rating agencies like KLD. Thus, we propose the following hypothesis:

Hypothesis (H1). *Continual CEP improvements are positively associated with aggregated environmental scores provided by sustainability rating agencies.*

3.2. Continual CEP improvements and CFP

The availability and quality of corporate environmental data has improved over the years, and CEP has been observable on the operational level for over a decade. Key initiators of environmental input and output data were mandatory reporting schemes from governmental agencies, such as the Toxic Release Inventory (TRI) and the Environmental Protection Agency's Greenhouse Gas Reporting Program (EPA GHGRP), as well as non-governmental agencies, such as the Carbon Disclosure Project (CDP). Furthermore, private data providers, including Bloomberg, MSCI, Sustainalytics, Thomson Reuters Asset4, and Trucost, gather and offer proprietary data on companies' environmental, social, and governance (ESG) activities for thousands of firms worldwide. These data providers offer a range of data points on an annual basis from disaggregated statistics on environmental performance (e.g., total energy use, carbon emissions, etc.).

As a result, studies have increasingly utilized disaggregated environmental measures (e.g., carbon emissions, toxic waste, etc.) as proxies for firms' CEP (e.g., Alvarez, 2012; Berchicci et al., 2017; Delmas et al., 2015; Misani & Pogutz, 2015). A focus on a single variable, e.g., carbon emissions, as a representation of a company's environmental impact may be seen as insufficient, as it only covers a single, often disparate snapshot of a company's overall environmental performance (Delmas et al., 2013; Misani & Pogutz, 2015). However, it has been argued that disaggregated measures can in certain cases provide greater indications of current environmental impacts than an entire environmental score (Chatterji et al., 2009). Depending on the measurement scale, i.e., either dichotomous data representing binary scales (yes/no) or continuous data with high variation (e.g., firms' resource use and emissions), we expect that continual CEP improvements – e.g., data about energy use, water use, carbon emissions, and water discharge – will increase the amount of explanatory power of the data. Therefore, the inclusion of continuous variables with high variation (e.g., carbon emissions) is considered more suitable to representing continual CEP improvements.

Previous studies have observed that improvements in environmental inputs are directly related to an increase in resource efficiency and productivity (Aragon-Correa & Sharma, 2003; Hart & Milstein, 2003; Hart & Dowell, 2011). Resource efficiency is a central term used in the assessment of environmental performance, as it seeks to decouple the direct relation between resource usage and environmental degradation (Christmann, 2000; Trumpp et al., 2015). The efficient use of key resources, i.e. energy and water, should reduce costs and enhance short- and long-term profitability due to gained operational efficiencies, which have a direct positive effect on firms' financial performance (Epstein & Roy, 1996; Salzmann et al., 2005).

The reduction of environmental outputs, including carbon emissions and water discharge, stipulate waste prevention as "the deviation of actual waste", as defined by Lenox and King (2002, p. 29). These authors find clear evidence that waste reduction contributes to a financial benefit, but they find no proof that companies profit from onsite treatment at the end of the pipe. Moreover, other research (Dutt & King, 2014) found that onsite treatment is associated with high investment costs and even increased waste amounts, as these amounts are correctly accounted for the first time. Since we focus on variables that signal the prevention of waste on the front end (e.g., reduced carbon emissions and lower amounts of water discharge; Lenox & King, 2002), we suggest that reductions of these environmental outputs will be positively associated with financial performance.

Furthermore, an ongoing inquiry as to whether and under what circumstances it pays to be green can be traced back to the 1970s (Grewatsch & Kleindienst, 2017). A good majority of recent studies find a consensus, that CEP does not have a negative association with CFP (Friede, Busch, & Bassen, 2016; Russo & Minto, 2012), and some studies can even establish a link that CEP positively influences CFP (e.g., Delmas et al., 2015; Eccles et al., 2014; Flammer, 2015; Hang et al., 2015). Thus, resource savings and waste prevention are found to increase internal profit margins (Porter & van der Linde, 1995), and have also been found to boost reputation and investor confidence, which would reflect positively on market-based figures (van Marrewijk, 2003). Therefore, the second hypothesis is framed as follows:

Hypothesis (H2). *Continual CEP improvements are positively associated with accounting-based and market-based CFP.*

The next section will describe the methodology, including how the samples were generated to test the hypotheses, and how the CEP and CFP measures were operationalized.

4. Methodology

4.1. Data collection

In order to estimate our panel regression models, we drew data from three prominent data providers: MSCI KLD, Thomson Reuters Asset4, and Compustat. These providers enabled us to connect longitudinal data on various environmental and financial aspects for the selected timeframe (2008 to 2016). The MSCI KLD environmental score is a "widely recognized benchmark for measuring the impact of […] environmental screening on investment portfolios" (Ortiz-de-Mandojana & Bansal, 2016, p. 37.) Thomson Reuters (2017) Asset4 offers hundreds of variables that portray various aspects of environmental performance, ranging from individual data points and indicator values to aggregate environmental pillar scores for more than 6,000 companies.

We searched for and selected four disaggregated environmental variables based on two criteria. For the first criteria, we searched for environmental variables that are common environmental indicators for continual improvements as indicated by international environmental/sustainability management norms, including the Global Reporting Initiative, the World Business Council for Sustainable Development, and two environmental management norms – ISO 14001 and the Eco-Management and Audit Scheme (EMAS).

Environmental indicators that reflect continual improvements include energy use (specified as the total direct and indirect energy used by a company in gigajoules), water use (specified as water withdrawal in cubic meters), carbon emissions (specified as total CO_2-equivalents for scope 1 and 2 emissions), the amount of water discharged (specified as emitted wastewater in cubic meters), and the amount of total waste (Brouwer & van Koppen, 2008). The second criteria was data availability. One variable in particular, waste, had missing data, especially when data was not available for the same company over multiple years. Thus, we omitted waste as a variable since it provided too little data to include in our regression models. Ultimately, we included four environmental indicators that had considerable observations for our panel data regressions: firms' total energy use, total water use, total CO_2-equivalent emissions (Scope 1 and 2), and total water discharge.

All financial data was gathered from Compustat. This included total annual sales, total assets, total equity, total liabilities, and market capitalization in order to calculate ROA and Tobin's q.

4.2. Data description

Depending on the respective models, our total sample ranged from roughly 500 to more than 3,000 firm-year observations in the years 2008–2016. Similar to previous studies (Trumpp & Guenther, 2017; Zhao & Murrell, 2016), we used the winsorizing technique at the top and the bottom 1st percentiles for all variables used in this study to account for potential outliers in the dataset. The following sections describe the dependent, independent, and control variables.

4.2.1. Dependent variables

KLD aggregates individual measures in various categories, including strengths and concerns for the areas of environment, community and society, employees and supply-chains, customers, and governance and ethics. Focusing on CEP, we selected the KLD environmental net score (i.e., environmental strengths minus environmental concerns). In addition, we included the individual KLD environmental strengths and KLD environmental concerns for a thorough analysis of environmental scores.

To test H2, we obtained financial data from Compustat, including net income as well as total assets, total equity, total liabilities, and market capitalization in order to calculate ROA and Tobin's q. These financial variables are frequently used in the literature researching the connection between CEP and CFP. For example, several prominent studies (Delmas et al., 2015; King & Lenox, 2002; Russo & Fouts, 1997) use ROA and Tobin's q exclusively to depict companies' financial performance. Thus, we include measures of profitability (i.e., accounting-based CFP portrayed as ROA) as well as market value (i.e., market-based CFP portrayed as

Tobin's q) to capture a firm's financial performance. ROA gives a manager, investor, or analyst an idea of how efficient a company's management is at using its assets to generate earnings. Tobin's q is calculated as the market value of a company divided by the replacement value of the firm's assets.

4.2.2. Independent variables

Our longitudinal database allowed us to capture the improvements of disaggregated CEP measures. For the disaggregated variables, we calculated the intensity of the respective environmental variables (e.g., total energy use divided by total annual sales), and then applied the dynamic factor as a year-to-year percentage change in intensity in order to arrive at our ΔCEP, which is similar to several previous studies (Alvarez, 2012; Lewandowski, 2017). We use intensities for these disaggregated environmental variables, as it makes it possible to compare the effects for companies of different sizes and in different industries. Since we use intensities, where a decrease in intensity is equivalent to an increase in environmental performance, we multiply our ΔCEP variables by (-1). Accordingly, a higher ΔCEP can be interpreted as a higher rate of improvement (Busch and Hoffmann, 2011). Following our variable treatment, ΔCEP will be time variant for all changes. Figure 1 below depicts our calculation to derive the disaggregated variables for measuring continual CEP improvements.

INSERT FIGURE 1 HERE

4.2.3. Control variables

There are a number of variables that are likely to be related to our dependent variables. Risk has been considered a source of firm-level heterogeneity in previous studies (e.g., Delmas et al., 2015; Waddock and Graves, 1997) and is included as a control variable, calculated as the debt to assets ratio. Additionally, we consider firm size to affect financial performance, and therefore include it in the form of a logarithm of total assets (*ln(assets)*). We also include R&D

intensity (i.e., R&D divided by sales) as a control, since it has been argued in the literature that distinctive technological capabilities can create value for firms (Barnett & Salomon, 2012; McWilliams & Siegel, 2000; Shrivastava, 1995), and thus can have an effect on CFP. Finally, we also included year dummies for our sample period to account for periodic effects.

4.3 Statistical approach

According to our two hypotheses, we analyzed the relationship of continual CEP improvements in two different ways: (i) its association with environmental scores, and (ii) its association with CFP. According to our first research question, we modeled the relationship between KLD and continual improvement of disaggregated environmental variables as follows:

$$KLD_{i,t} = \alpha + \beta \Delta CEP_{i,t} + \pi Controls_{i,t} + Year_{i,t} + (\Omega Individual_i + \mu_{i,t})$$

Where time is denoted by t and firms are denoted by i. KLD is our dependent variable, and *ΔCEP* is our explanatory variable of interest, which includes *ΔEnergy, ΔWater,* ΔCO_2, and *ΔWaterDischarge*. Controls contain explanatory variables that might affect the financial performance of firms, including *Risk, Size*, and *R&D intensity*. *Individual* represents firm fixed effects, which control for firm unobserved heterogeneity that captures any time-invariant firm characteristics. In all models, we use firm as well as *year fixed effects* to denote all company and time effects that capture common shocks, like financial crises, changes in government policy, or other systematic macroeconomic shocks that affect the financial performance of all firms. μ is the error term that captures all other omitted factors.

Next, we considered the association of continual improvements of disaggregated environmental variables with CFP and modified the econometric model accordingly:

$$CFP_{i,t} = \alpha + \beta \Delta CEP_{i,t} + \pi Controls_{i,t} + Year_{i,t} + (\Omega Individual_i + \mu_{i,t})$$

Again, time is denoted by t and firms are denoted by i. Here, CFP is our dependent variable, including *ROA* and *Tobin's q.* As above, the *ΔCEP* included *ΔEnergy, ΔWater,* ΔCO_2, and *ΔWaterDischarge*. For the controls, we include *Risk, Size,* and *R&D Intensity*.

For both basic models, we started with a pooled ordinary least squares (OLS) model, which is consistent and efficient if individual heterogeneity is not expected (or *Individual* only contains a constant term). In contrast, both random effects and fixed effects assume the existence of unobserved individual heterogeneity (*Individual*), with fixed effects models allowing for correlation between *Individual* and *ΔCEP* or *Control*, whereas random effects models do not allow for this correlation. Thus, if *Individual* is uncorrelated with *ΔCEP* or *Control*, random effects will be more efficient, while if *Individual* is correlated with *ΔCEP* or *Control*, random effects will be biased but fixed effects will be consistent (c.f. Greene, 2008).

First, we conducted the Breusch-Pagan Lagrange Multiplier (LM) test to decide between a random effects or pooled OLS regression. The LM test has a null hypothesis of zero variances across individuals. To decide between random and fixed effects, we ran a Hausman test with null hypothesis of µit not correlated with other control variables (*Δ-CEP* or Control). Our test results show that fixed effects is the most appropriate econometric model. We use standard errors that are robust to heteroscedasticity and serial correlation in all fixed effects models.

5. Results

5.1. Descriptive statistics

Table 1 presents the descriptive statistics for the underlying set of variables. All *KLD variables* are characterized by positive mean values. Thus, the positive mean values indicate that environmental strengths compensate for environmental concerns. For all *ΔCEP variables,* the mean values are negative. Based on our variable treatment (i.e., multiplying the intensities by -1), a negative value for the ΔCEP variables implies an increase of resource intensities for the entire sample, and an overall decrease in CEP. In other words, environmental performance (e.g., higher use of energy per sales) is worsening on average. Thus, no systematic improvements could be derived from the mean values.

Additionally, all variables appear to be within a reasonable range of variation. For ROA, we find a mean value of 4.5%, while the mean value of Tobin's q is greater than 1. This indicates that, on average, the market value of companies is higher than the recorded value of assets.

INSERT TABLE 1 HERE

5.2. Main results

For H1, we created four sets consisting of three models for each *Δ*CEP variable. The first three models reveal the association of *ΔEnergy* with the KLD environmental net score, KLD environmental strengths, and KLD environmental concerns. For *ΔEnergy*, we find moderate (b = -0.630, p = 0.147, Model 1 in Table 2) and strong (b = 0.454, p = 0.016, Model 3 in Table 2) support for the result that improvements in energy intensities are reflected by the KLD environmental net score and KLD environmental concerns; however, not as we might have expected. Rather, improvements in energy intensities reduce the environmental net score and increase environmental concerns respectively.

We obtained similar results for the other *ΔCEP variables.* For *ΔWater* (b = 0.292, p = 0.087, Model 6 in Table 2) and *ΔWaterDischarge* (b = 0.513, p = 0.087, Model 12 in Table 2), we also find a positive association with environmental concerns. Additionally, we observed a negative association between *Δ*CO_2 and the KLD environmental net score (b = -0.760, p = 0.011, Model 7 in Table 2) as well as between *Δ*CO_2 and environmental strengths (b = -0.532, p = 0.055, Model 8 in Table 2). In line with our expectations, our control variables follow similar patterns as indicated by previous research, while we find different effect sizes between *ΔCEP input* and *output* models.

Consequentially, we reject H1 in acknowledgment of these surprising results. We found mostly negative or no significant associations between CEP improvements and KLD environmental scores, including environmental strengths and weaknesses. Thus, continual CEP improvements of disaggregated environmental indicators are *not* captured by KLD environmental scores, and, in some cases, they even worsen a firm's environmental score.

INSERT TABLE 2 HERE

With respect to H2, we tested the association of continual CEP improvements (*ΔEnergy, ΔWater, Δ*CO_2, and *ΔWaterDischarge)* on CFP. Table 3 presents the results for two models per ΔCEP indicator, one model for accounting-based CFP, i.e., ROA, and another for market-based CFP, i.e., Tobin's q. In total, we regressed our independent variables in eight different panel data models. Additionally, all models were controlled for year and firm-fixed effects, as well as for size, risk, and R&D intensity. We find highly significant, positive associations with both accounting-based as well as market-based CFP indicators for all *ΔCEP variables* (at the $\rho<0.01$ level).

Following our variable design, we are also able to analyze the magnitude of the individual impacts. For ROA, we find the greatest effects resulting from changes in *Δ-WaterDischarge* and the smallest resulting from changes in *Δ-Energy*. Here, for example, a one percent increase in *WaterDischarge* intensity is associated with an increase in ROA by 0.0299 percent points. Also for Tobin's Q, a one percent increase in *WaterDischarge* has the largest effect, resulting into an improvement of 0.208 for Tobin's Q. On the other side, improvements in CO_2 intensity have the smallest effect on Tobin's Q (b=0.191, p=0.000 Model 6 in Table3). Thus, the results strongly support H2, as continual CEP improvements are positively associated with CFP. As expected, the results for our control variables remain quite stable over the corresponding panel models.

INSERT TABLE 3 HERE

5.3. Robustness Checks

We performed several robustness checks to test, and even reduce, potential issues stemming from our CEP measures, especially the issues of data heterogeneity and endogeneity (Surroca et al., 2010). First, we substituted the KLD environmental score for another established environmental score, namely the Asset4 environmental pillar score, and reassessed H1. The Asset4 environmental pillar score has been used in recent CEP studies, and it has been claimed as a robust measure to capture the overall CEP of firms by including both managerial and operational issues (Ioannou and Serafeim, 2012; Cheng et al., 2014; Misani and Pogutz, 2015; Trumpp et al., 2015). According to Thomson Reuters (2017, p. 12), the environmental pillar combines the areas of resource use, emissions, and environmental innovation, which reflects "a company's performance and capacity to reduce the use of materials, energy or water…[as well as its]…commitment and effectiveness towards reducing environmental emissions in the production and operational processes".

INSERT TABLE 4 HERE

The results reveal that utilizing the Asset4 environmental pillar score did not change our findings for H1. In fact, here we do not find any significant relationship between our *Δ-CEP variables* and the Asset4 environmental pillar score.

Subsequently, we addressed possible endogeneity issues (i.e., reverse causality) to reassess the results corresponding to H2. For this, we applied additional changes to check the robustness of our research models. At this point, we lagged dependent variables, i.e., the CFP variables ROA and Tobin's q, to reduce influences resulting from endogeneity (Surroca et al., 2010). Thus, we can argue that CEP improvements are not driven by significant changes in CFP in a specific year. The inclusion of these lagged dependent variables did not affect the previous results for our second hypotheses, as all models provided the same significant and positive coefficients.

Furthermore, we introduced a new variable to measure for continual CEP improvements over multiple years, which we have termed *cumulative effect*. In other words, we checked how often a firm consistently improves in order to capture a firm's track record of continual CEP improvements over the years. The calculation of the variable is based on accumulative years that all firms experience an improvement in the years tested. For example, if a company improved in consecutive years 2008–2010 (three years), but failed to improve between 2011 and 2012, and then improved again in consecutive years 2013–2016 (four years), we would consider accumulated improvements to equal seven in the final calculation.

INSERT TABLE 5 HERE

The results in Table 5 support our previous findings. First, we still find highly significant positive relationships between all *ΔCEP* and CFP measures in all additional research models based on the cumulative effect. This reinforces H2, which assumes that improvements of operational environmental indicators positively influence a CFP. We find strong support that cumulative improvements in *ΔEnergy* (b = 0.002, p = 0.070, Model 1 and b = 0.045, p = 0.000, Model 2 in Table 5) and *Δ*CO_2 (b = 0.003, p = 0.005, Model 5 and b = 0.038, p = 0.000, Model 6 in Table 5) positively influence ROA and Tobin's q. Thus, companies that improve their energy and CO_2 intensities over consecutive years will be more profitable and financially rewarded. For *ΔWater* and *ΔWaterDischarge*, we mainly find neutral and insignificant results; however, we do not find negative results. Only for subsequent improvements in *ΔWater* and Tobin's q do we find a positive relationship (b = 0.047, p = 0.000, Model 4 in Table 5). Thus, the cumulative improvements concerning the management of a company's water cycles are only reflected in CFP to a limited extent. In sum, we find that continual CEP improvements are rewarded in general, and in the case of *ΔEnergy* and *Δ*CO_2 we detect a positive cumulative effect. The implications of these findings for future research and practice will be discussed in the next section.

6. Discussion and Conclusion

In this article, we predicted that continual CEP improvements would be reflected in environmental scores from sustainability rating agencies (i.e., KLD and Asset4 environmental scores) and that such improvements are positively associated with enhanced CFP. On the one hand, we found that continual CEP improvements are not adequately reflected in aggregated environmental scores provided by sustainability rating agencies. On the other hand, we found strong support for the prediction that multiple operational measures of continual CEP improvements are positively associated with the CFP measures. The results stress that continual CEP improvements are not just an important capability to achieve environmental reduction targets, but are also relevant from a materiality point of view.

Our study contributes to the strategic management literature in two main regards. As our first contribution, we show that continual CEP improvements are not adequately reflected in environmental scores, including KLD and Asset4, which is surprising. Recent studies have investigated issues regarding the validity and consistency of data provided by sustainability rating agencies (Berg et al., 2019; Chatterji et al., 2016; Delmas et al., 2013; Eccles et al., 2019). We are able to contribute to this literature from a novel perspective by revealing that continual CEP improvements are not adequately reflected in environmental scores. We should emphasize that sustainability rating agencies capture a variety of aspects when assessing CEP; however, continual improvements turn out to be obscure in the mix. The detected inconsistencies between continual CEP improvements and aggregated environmental scores are a starting place for further understanding of this relationship and further debate about the role of sustainability rating agencies.

As we cannot confirm the association between continual CEP improvements and aggregated environmental scores, we searched for possible explanations for this discord beyond the results. One potential reason for insignificant results may be due to the methodologies, i.e., the data collection and measurements, of aggregated environmental scores. Environmental scores, such as KLD, are based on multiple environmental criteria. Scores do incorporate operational indicators, but they also include issues dealing with environmental strategies, policies, adoption and adherence to existing norms, and a number of other environmental initiatives. Variables concerned with environmental management policies, norms, and systems, which are reflected in these environmental scores, do not necessarily capture continual improvements, as these measures reflect rather static developments in CEP (i.e., whether a company has an environmental policy or not), which, when aggregated, may limit other areas of greater variation. Thus, KLD environmental scores contain rather static properties of CEP, which may weaken the effects of continual CEP improvements based on an operational outcomes.

Additionally, KLD experienced a disruption in assessment criteria in 2011, which led to a change in data collection and measurement (Eccles et al., 2019). Thus, the firm-year observations in the years 2008– 2016 could be severely compromised, especially if we had not addressed the problem with yearly fixed effects and further robustness testing. Thus, we included the Asset4 Environmental Pillar Score, but the results did not change. This might be explained by other measurement techniques, including industrial benchmarks and weighted averages. Such weighted scores should be included in regional and industrial considerations, and this may be extremely useful for comparisons between companies (i.e., in identical industries); however, such measurements may dilute idiosyncrasies of individual firms, which most likely include continual CEP improvements.

Furthermore, the resulting negative coefficients in H1, e.g., ΔCO_2 (Table 2), indicate that companies may be improving their aggregated environmental score under the scrutiny of rating agencies, while their operational performance declines. Apparently, the improvement in other environmental aspects may be substantial, i.e., more environmental policies and less environmental concerns, which overcompensates for the worsening environmental operational performance. For example, a multinational delivery company (intentionally anonymous) improved its KLD environmental net score from 1 to 5 between the years 2006–2013. During the same time, however, this company's absolute carbon emissions rose from 7.73 million metric tons of CO_2-equivalent emissions (CO_2e) to 12.6 million metric tons of CO_2e. Converting carbon emissions into an intensity rate (i.e., CO_2e divided by sales), we found that the company's CO_2e intensity worsened by 50% over this period. Precisely, the environmental score of this company improved tremendously, while the company's carbon performance declined in terms of absolute and relative performance. Thus, aggregated environmental scores may often fail to capture continual CEP improvements adequately.

Our second contribution to the literature relates to organizational capabilities. Previous literature has argued that continual improvements should be considered as a valuable capability (Aragón-Corrrea & Sharma, 2003; Hart, 1995; Hart & Milstein, 1993; Russo and Fouts, 1997; Sharma & Vredenburg, 1998). Our results imply that continual CEP improvements have a significant positive effect on profitability. This empirical outcome confirms early research that mostly theorized about continual CEP improvements in the context of operational efficiencies (Hart, 1995; Hart & Milstein, 2003). If firms can realize continual CEP improvements as a capability to increase operational efficiencies, they should be able to improve profitability and investor confidence for the long-term (Dowell & Muthulingam, 2017; Ortiz-de-Mandojana & Bansal, 2016).

When evaluating the impact of continual CEP improvements, we can confirm that recurring improvements, in terms of reduced energy use and carbon emissions, will lead to both short-term and long-term financial benefits. This finding contests previous findings that continual improvements of emission reductions will be associated with lower short-term financial performance, especially in times of regulatory uncertainty (Alvarez, 2012; Delmas et al., 2015). On the contrary, our finding demonstrates that continual improvements to companies' carbon performance positively influence short-term profitability, represented as ROA. As such, we conclude that continual CEP improvements in terms of energy savings, carbon reductions, and pollution prevention are a reflection of good management practices (King & Lenox, 2002).

Disaggregated environmental indicators related to operational processes and environmental impacts, in particular energy use, water use, CO_2e emissions, and water discharge, are very fitting as measures of continual improvement. We would like to emphasize that CEP is more than the culmination of environmental outcomes and impacts; it is how these outcomes are related back to corporate strategy and policies that encourages improvement (Jiang & Bansal, 2003; Link & Nevah, 2006). Thus, managers, investors, and academics should not rely merely on aggregated environmental scores provided by sustainability rating agencies, but should consider continual CEP improvements and environmental scores in an integrated and holistic manner.

We derive several implications for practitioners. First, we highlight that existing environmental scores, i.e., KLD and Asset4, capture the level of CEP but not necessarily the rate of improvement in individual firms. Often investors rely on so-called “best-in-class” approaches to assess CEP, which includes positive and negative screenings (EUROSIF, 2016). On the one hand, these abovementioned scores prompt greater transparency of non-financial issues and promote commitment to sustainability principles within companies. On the other hand, however, these approaches and ratings typically measure CEP as highly aggregated and mostly as a static phenomenon, where the scope of environmental practices are an aggregate of binary metrics or categorical measures with low variation. The “best-in-class” approach has its own drawbacks, as companies with previously positive sustainability ratings often remain in these indices for long durations, which also implies that companies that recently began improving CEP on the operational level may not qualify to improve scores.

Second, continual CEP improvements on the operational level contain financially material information. Currently, an unresolved debate exists between practitioners and both corporate managers and financial investors concerning which environmental aspects should be considered financially material (Berg et al., 2019; Eccles et al., 2019). Managers and investors should concentrate not only on the level of CEP achieved, but also on the rate of improvement between years. Thus, firms should monitor and actively address mitigation programs to reduce environmental footprints. From a holistic perspective, a new range of indices could measure and promote *best-in-progress*, meaning those firms that substantially improve compared to others in their industry, region, and overall. We do not suggest *best-in-progress* should replace the existing *best-in-class* approach, but they could be used together in a complementary fashion.

Third, managers and investors must carefully scrutinize this information in order to gain a holistic view of all different aspects of CEP, and not just those stemming from environmental scores provided by sustainability rating agencies. We are encouraged to find consistent positive results for H2, implying that continual CEP improvements of all disaggregated measures, including energy use, water use, carbon emissions, and water discharge, have a significant, positive effect on both firm profitability (ROA) and market value (Tobin's q) of firms. For example, significant reductions of carbon emissions are required in order to achieve the science-based targets well below 2°C temperature rise (Hulme, 2016). By considering continual CEP improvements as financially material, investors could strengthen the effort to proactively finance the UN Agenda 2030 Sustainable Development Goals by opening the door to a different set of environmental indicators, including climate mitigation, biodiversity gains, and resource efficiency (Howard-Grenville et al., 2017). Following this view, companies currently considered environmental laggards could possibly be rewarded for vast environmental improvements, rather than limiting the focus to the top-rated companies. This article addresses this issue by suggesting a new and complementary way to capture continual environmental improvements and investigate their influence on CFP.

Finally, we recognize several limitations in our study. First, we realize that other CEP variables exist, which could qualify as indicators for continual CEP improvement. Brouwer and van Koopen (2008) listed multiple indicators in international environmental management frameworks, including operational indicators that we selected (e.g., energy use, water use, carbon emissions, and water discharge). However, they also provided further operational indicators (e.g., total waste, recollection of waste) as well as management indicators (e.g., total environmental investments, amount of environmental education, etc.). While the exclusion of additional operational indicators was due to restricted data on these items, we purposefully omitted environmental management indicators, as we strongly focus on the operational side of CEP for continual improvements. Future studies may want to investigate further CEP variables, to test the suitability and effects on various measures (e.g., CFP measures).

Additionally, we recognize the inclusion of KLD environmental scores as a limitation. KLD environmental scores have been wrought with debate in previous academic research. Previous studies have found that KLD strengths and concerns should not be combined, and numerous studies are already incorporating datasets from other sustainability rating agencies and data providers (e.g., Thomson Reuters ASSET4; Cheng et al., 2014; Ioannou & Serafeim, 2012; Misani & Pogutz, 2015), which allow for much more fine-grained and accurate measurements of CEP. In this article, we have aimed to understand the extent of this limitation through the disaggregation of KLD scores (strengths and concerns) and the inclusion of robustness checks. Thus, we have confidence that the results presented in this article are still acceptable.

Future research might also consider further sustainability-related criteria deemed salient for improvements. For example, social variables could include health and safety issues, working conditions for employees, and improved standards in supply chains (Ruf, Muralidhar, Brown, Janney, & Paul, 2001). The empirical validation of further sustainability variables according to our approach depends largely on the availability of data, i.e., sufficient observations for many companies over multiple years. Nevertheless, a broader collection of social and sustainability performance data would provide a wider understanding of companies' continual improvements in sustainability performance.

References

Alvarez, I.G. (2012). Impact of CO_2 Emission Variation on Firm Performance. *Business Strategy and the Environment*, *21*(7), 435–454.

Anand, G., Ward, P. T., Tatikonda, M. V., & Schilling, D. A. (2009). Dynamic capabilities through continuous improvement infrastructure. *Journal of Operations Management*, *27*(6), 444-461.

Bansal, P., & Roth, K. (2000). Why companies go green: A model of ecological responsiveness. *Academy of Management Journal*, *43*(4), 717-736.

Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, *33*(11), 1304-1320.

Berchicci, L., Dowell, G., & King, A. A. (2012). Environmental capabilities and corporate strategy: Exploring acquisitions among US manufacturing firms. *Strategic Management Journal*, *33*(9), 1053-1071.

Berchicci, L., Dowell, G., & King, A. A. (2017). Environmental performance and the market for corporate assets. *Strategic Management Journal*, *38*(12), 2444-2464.

Berg, F., Koelbel, J. F., & Rigobon, R. (2020). Aggregate confusion: the divergence of ESG ratings. *Available at SSRN:*https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=3438533.

Brouwer, M. A., & van Koppen, C. K. (2008). The soul of the machine: continual improvement in ISO 14001. *Journal of Cleaner Production*, *16*(4), 450-457.

Busch, T., & Hoffmann, V. H. (2011). How hot is your bottom line? Linking carbon and financial performance. *Business & Society*, *50*(2), 233-265.

Buysse, K., & Verbeke, A. (2003). Proactive environmental strategies: A stakeholder management perspective. *Strategic Management Journal*, *24*(5), 453-470.

Certo, S. T., Withers, M. C., & Semadeni, M. (2017). A tale of two effects: Using longitudinal data to compare within-and between-firm effects. *Strategic Management Journal*, *38*(7), 1536-1556.

Chatterji, A., & Levine, D. (2006). Breaking down the wall of codes: Evaluating non-financial performance measurement. *California Management Review*, *48*(2), 29-51.

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, *18*(1), 125-169.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, *37*(8), 1597-1614.

Chen, C. M., & Delmas, M. (2011). Measuring corporate social performance: An efficiency perspective. *Production and Operations Management*, *20*(6), 789-804.

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

Christmann, P. (2000). Effects of "best practices" of environmental management on cost advantage: The role of complementary assets. *Academy of Management Journal*, *43*(4), 663-680.

Darnall, N., & Edwards Jr, D. (2006). Predicting the cost of environmental management system adoption: the role of capabilities, resources and ownership structure. *Strategic Management Journal*, *27*(4), 301-320.

Delmas, M. A., Etzion, D., & Nairn-Birch, N. (2013). Triangulating environmental performance: What do corporate social responsibility ratings really capture?. *Academy of Management Perspectives*, *27*(3), 255-267.

Delmas, M. A., Nairn-Birch, N., & Lim, J. (2015). Dynamics of environmental and financial performance: The case of greenhouse gas emissions. *Organization & Environment*, *28*(4), 374-393.

Delmas, M. A., & Toffel, M. W. (2008). Organizational responses to environmental demands: Opening the black box. *Strategic Management Journal*, *29*(10), 1027-1055.

Deming WE. 1986. *Out of the Crisis*. Cambridge, MA: MIT Press.

Dixon-Fowler, H. R., Slater, D. J., Johnson, J. L., Ellstrand, A. E., & Romi, A. M. (2013). Beyond "does it pay to be green?" A meta-analysis of moderators of the CEP–CFP relationship. *Journal of Business Ethics*, *112*(2), 353-366.

Dowell, G. W., & Muthulingam, S. (2017). Will firms go green if it pays? T he impact of disruption, cost, and external factors on the adoption of environmental initiatives. *Strategic Management Journal*, *38*(6), 1287-1304.

Dutt, N., & King, A. A. (2014). The judgment of garbage: End-of-pipe treatment and waste reduction. *Management Science*, *60*(7), 1812-1828.

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835-2857.

Eccles, R. G., Lee, L. E., & Stroehle, J. C. (2019). The social origins of ESG: An analysis of Innovest and KLD. *Organization & Environment*, Online First. https://doi.org/10.1177/1086026619888994

Erhemjamts, O., Li, Q., & Venkateswaran, A. (2013). Corporate social responsibility and its impact on firms' investment policy, organizational structure, and performance. *Journal of Business Ethics*, *118*(2), 395-412.

Elsayed, K., & Paton, D. (2005). The impact of environmental performance on firm performance: static and dynamic panel data evidence. *Structural Change and Economic Dynamics*, *16*(3), 395-412.

Endrikat, J., Guenther, E., & Hoppe, H. (2014). Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal*, *32*(5), 735-751.

Epstein, M. J., & Roy, M. J. (2003). Making the business case for sustainability: linking social and environmental actions to financial performance. *Journal of Corporate Citizenship,* (9), 79-96.

Etzion, D. (2007). Research on organizations and the natural environment, 1992-present: A review. *Journal of Management*, *33*(4), 637-664.

EUROSIF (2016). European Social Investment Forum SRI Study 2016. Retrieved from https://www.eurosif.org/sri-study-2016/

Flammer, C. (2015). Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. *Management Science*, *61*(11), 2549-2568.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210-233.

Greene, W.H. (2008). The econometric approach to efficiency analysis. In H.O. Fried, C.A. Knox Lovell, & S.S. Schmidt (Eds.), *The measurement of productive efficiency and productivity growth*: 92–250. Oxford, UK: Oxford University Press.

Grewatsch, S., & Kleindienst, I. (2017). When does it pay to be good? Moderators and mediators in the corporate sustainability–corporate financial performance relationship: A critical review. *Journal of Business Ethics*, *145*(2), 383-416.

Hang, M., Geyer-Klingeberg, J., & Rathgeber, A. W. (2019). It is merely a matter of time: A meta-analysis of the causality between environmental performance and financial performance. *Business Strategy and the Environment*, *28*(2), 257-273.

Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of Management Review*, *20*(4), 986-1014.

Hart, S. L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business strategy and the Environment*, *5*(1), 30-37.

Hart, S. L., & Dowell, G. (2011). Invited editorial: a natural-resource-based view of the firm: fifteen years after. *Journal of Management*, *37*(5), 1464-1479.

Hart, S. L., & Milstein, M. B. (2003). Creating sustainable value. *Academy of Management Perspectives*, *17*(2), 56-67.

Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal*, *24*(10), 997-1010.

Helfat, C. E., & Winter, S. G. (2011). Untangling dynamic and operational capabilities: Strategy for the (N) ever-changing world. *Strategic Management Journal*, *32*(11), 1243-1250.

Horváthová, E. (2012). The impact of environmental performance on firm performance: Short-term costs and long-term benefits? *Ecological Economics*, *84*, 91-97.

Howard-Grenville, J., Davis, J., Dyllick, T., Joshi, A., Miller, C., Thau, S., & Tsui, A. S. (2017). Sustainable development for a better world: Contributions of leadership, management and organizations: Submission deadline: July 1 to July 30, 2018. *Academy of Management Discoveries*, *3*(1), 107-110.

Hulme, M. (2016). 1.5 C and climate research after the Paris Agreement. *Nature Climate Change*, *6*(3), 222-224.

Ilinitch, A. Y., Soderstrom, N. S., & Thomas, T. E. (1998). Measuring corporate environmental performance. *Journal of Accounting and Public Policy*, *17*(4-5), 383-408.

International Organization for Standardization (ISO) (2013). ISO 14031: 2013. *Environmental management—environmental performance evaluation—guidelines*. Geneva: ISO.

International Organization for Standardization (ISO) (2015). ISO 14001: 2015. *Environmental management systems: Requirements with guidance for use*. Geneva: ISO.

Ioannou, I., & Serafeim, G. (2012). What drives corporate social performance? The role of nation-level institutions. *Journal of International Business Studies*, *43*(9), 834-864.

Ioannou, I., & Serafeim, G. (2015). The impact of corporate social responsibility on investment recommendations: Analysts' perceptions and shifting institutional logics. *Strategic Management Journal*, *36*(7), 1053-1081.

Jiang, R. J., & Bansal, P. (2003). Seeing the need for ISO 14001. *Journal of Management Studies*, *40*(4), 1047-1067.

Jones, T. M. (1995). Instrumental stakeholder theory: A synthesis of ethics and economics. *Academy of Management Review*, *20*(2), 404-437.

King, A., & Lenox, M. (2002). Exploring the locus of profitable pollution reduction. *Management Science*, *48*(2), 289-299.

Ledford GE. 2015. Continuous improvement. In C.L. Cooper, P.C. Flood, & Y. Freeney (Eds.), *Wiley Encyclopedia of Management (Volume 11): Organizational Behavior* (Online only). https://doi.org/10.1002/9781118785317.weom110130

Lewandowski, S. (2017). Corporate carbon and financial performance: The role of emission reductions. *Business Strategy and the Environment*, *26*(8), 1196-1211.

Link, S., & Naveh, E. (2006). Standardization and discretion: does the environmental standard ISO 14001 lead to performance benefits?. *IEEE Transactions on Engineering Management*, *53*(4), 508-519.

López, M. V., 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), 285-300.

Lundgren, T., & Zhou, W. (2017). Firm performance and the role of environmental management. *Journal of Environmental Management*, *203*, 330-341.

Mattingly, J. E. (2017). Corporate social performance: A review of empirical research examining the corporation–society relationship using Kinder, Lydenberg, Domini social ratings data. *Business & Society*, *56*(6), 796-839.

Mattingly, J. E., & Berman, S. L. (2006). Measurement of corporate social action: Discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business & Society*, *45*(1), 20-46.

McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, *31*(4), 854-872.

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

Misani, N., & Pogutz, S. (2015). Unraveling the effects of environmental outcomes and processes on financial performance: A non-linear approach. *Ecological Economics*, *109*, 150-160.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, *24*(3), 403-441.

Ortiz-de-Mandojana, N., & Bansal, P. (2016). The long-term benefits of organizational resilience through sustainable business practices. *Strategic Management Journal*, *37*(8), 1615-1631.

Perrault, E., & Quinn, M. A. (2018). What have firms been doing? Exploring what KLD data report about firms' corporate social performance in the period 2000-2010. *Business & Society*, *57*(5), 890-928.

Porter, M. E., & Van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *Journal of Economic Perspectives*, *9*(4), 97-118.

Ruf, B. M., Muralidhar, K., Brown, R. M., Janney, J. J., & Paul, K. (2001). An empirical investigation of the relationship between change in corporate social performance and financial performance: A stakeholder theory perspective. *Journal of Business Ethics*, *32*(2), 143-156.

Russo, M. V., & Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal*, *40*(3), 534-559.

Russo, M. V., & Minto, A. (2012). Competitive strategy and the environment: A field of inquiry emerges. In P. Bansal & A.J. Hoffman (Eds), *The Oxford Handbook of Business and the Natural Environment*: 29–49. Oxford, UK: Oxford University Press.

Salzmann, O., Ionescu-Somers, A., & Steger, U. (2005). The business case for corporate sustainability: literature review and research options. *European Management Journal, 23*(1), 27-36.

Schultze, W., & Trommer, R. (2012). The concept of environmental performance and its measurement in empirical studies. *Journal of Management Control*, *22*(4), 375-412.

Semenova, N., & Hassel, L. G. (2015). On the validity of environmental performance metrics. *Journal of Business Ethics*, *132*(2), 249-258.

Sharma, S., & Vredenburg, H. (1998). Proactive corporate environmental strategy and the development of competitively valuable organizational capabilities. *Strategic Management Journal*, *19*(8), 729-753.

Short, J. C., McKenny, A. F., Ketchen, D. J., Snow, C. C., & Hult, G. T. M. (2016). An empirical examination of firm, industry, and temporal effects on corporate social performance. *Business & Society*, *55*(8), 1122-1156.

Shrivastava, P. (1995). Environmental technologies and competitive advantage. *Strategic Management Journal, 16*(S1), 183-200.

Simpson, W. G., & Kohers, T. (2002). The link between corporate social and financial performance: Evidence from the banking industry. *Journal of Business Ethics*, *35*(2), 97-109.

Slawinski, N., & Bansal, P. (2015). Short on time: Intertemporal tensions in business sustainability. *Organization Science*, *26*(2), 531-549.

Suddaby, R. (2010). Editor's comments: Construct clarity in theories of management and organization. Academy of Management Review, 35(3), 346–357.

Surroca J., Tribó, J.A., & Waddock S. (2010). Corporate responsibility and financial performance: The role of intangible resources. *Strategic Management Journal*, 31(5), 463–490.

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, *28*(13), 1319-1350.

Thomson Reuters. (2017). Thomson Reuters ESG Scores. Retrieved from https://financial.thomsonreuters.com/ content/dam/openweb/documents/pdf/financial/esg-scores-methodology.pdf

Trumpp, C., Endrikat, J., Zopf, C., & Guenther, E. (2015). Definition, conceptualization, and measurement of corporate environmental performance: A critical examination of a multidimensional construct. *Journal of Business Ethics*, *126*(2), 185-204.

Trumpp, C., & Guenther, T. (2017). Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. *Business Strategy and the Environment*, *26*(1), 49-68.

Tsai, K. H., & Wang, J. C. (2005). Does R&D performance decline with firm size?—A re-examination in terms of elasticity. *Research Policy*, *34*(6), 966-976.

Van Marrewijk, M. (2003). Concepts and definitions of CSR and corporate sustainability: Between agency and communion. *Journal of Business Ethics*, *44*(2-3), 95-105.

Voss, C.A. (2005). Paradigms of manufacturing strategy re-visited. *International Journal of Operations & Production Management*, *25*(12), 1223–1227.

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

Walls, J. L., Phan, P. H., & Berrone, P. (2011). Measuring environmental strategy: Construct development, reliability, and validity. *Business & Society*, *50*(1), 71-115.

Yadav, P. L., Han, S. H., & Kim, H. (2017). Sustaining competitive advantage through corporate environmental performance. *Business Strategy and the Environment*, *26*(3), 345-357.

Zhao, X., & Murrell, A. J. (2016). Revisiting the corporate social performance-financial performance link: A replication of Waddock and Graves. *Strategic Management Journal*, *37*(11), 2378-2388.

Zollo, M., Cennamo, C., & Neumann, K. (2013). Beyond what and why: Understanding organizational evolution towards sustainable enterprise models. *Organization & Environment*, *26*(3), 241-259.

Figure 1. Calculation for continual CEP improvements (ΔCEP)

$$\Delta\, CEP \;=\; \frac{\dfrac{CEP}{Sales_t} - \dfrac{CEP}{Sales_{t-1}}}{\dfrac{CEP}{Sales_{t-1}}}$$

Table 1. Descriptive Statistics and Correlations

Variables	Mean	Std Dev	Min	Max	ROA	Tobin's q	Δ-KLD ENV	Δ-KLD Env Str	Δ-KLD Env Con	Δ-Energy	Δ-Water	Δ-CO2	Δ-Water Discharge
ROA	.0447	.0489	-.1017	.1744	1								
Tobin's q	1.592	.7146	.8395	4.2838	0.5145*	1							
KLD ENV	.1793	.9886	-5	6	0.0479*	0.0602*	1						
KLD Env Str	.4470	.9136	0	6	0.0663*	0.0149	0.7260*	1					
KLD Env Con	.2676	.7075	0	5	0.0188	-0.0651*	-0.4598*	0.2769*	1				
ΔEnergy	-.0198	.1650	-.6436	.3351	0.1110*	0.0760*	-0.0205	0.0040	0.0368	1			
ΔWater	-.0117	.1752	-.6567	.3504	0.1158*	0.0649*	-0.0156	-0.0163	0.0012	0.5306*	1		
ΔCO_2	-.0094	.1640	-.5734	.3321	0.1297*	0.0789*	0.0024	0.0124	0.0129	0.6325*	0.4813*	1	
ΔWater Discharge	-.0212	.1916	-.7137	.3723	0.1448*	0.0894*	-0.0587	-0.0672	0.0008	0.4646*	0.6552*	0.4195*	1

Table 2. Regression results for continual CEP improvements and KLD scores

VARIABLES	(1) KLDenv	(2) ENVstr	(3) ENVcon	(4) KLDenv	(5) ENVstr	(6) ENVcon	(7) KLDenv	(8) ENVstr	(9) ENVcon	(10) KLDenv	(11) ENVstr	(12) ENVcon
ΔENERGY	-0.630	-0.176	0.454									
	(0.147)	(0.635)	(0.016)									
ΔWATER				-0.459	-0.167	0.292						
				(0.301)	(0.658)	(0.087)						
ΔCO_2							-0.760	-0.532	0.228			
							(0.011)	(0.055)	(0.177)			
ΔWATERDISCHARGE										-0.950	-0.437	0.513
										(0.119)	(0.307)	(0.087)
ROA	-5.284	-3.777	1.508	-3.739	-2.733	1.007	-1.981	-1.589	0.392	-1.152	-0.914	0.238
	(0.003)	(0.026)	(0.123)	(0.039)	(0.158)	(0.384)	(0.290)	(0.291)	(0.674)	(0.789)	(0.727)	(0.935)
Size	0.203	0.574	0.371	0.0238	0.368	0.344	0.354	0.822	0.468	0.248	0.326	0.0779
	(0.541)	(0.077)	(0.131)	(0.939)	(0.337)	(0.189)	(0.273)	(0.005)	(0.099)	(0.707)	(0.419)	(0.815)
Risk	-2.656	-2.387	0.268	-1.979	-1.956	0.0233	-0.850	-0.925	-0.0759	-2.896	-3.114	-0.218
	(0.026)	(0.038)	(0.677)	(0.131)	(0.105)	(0.973)	(0.529)	(0.428)	(0.902)	(0.149)	(0.020)	(0.841)
R&D Intensity	6.511	4.880	-1.631	4.592	2.681	-1.911	3.959	2.940	-1.019	8.957	9.903	0.946
	(0.059)	(0.065)	(0.242)	(0.174)	(0.347)	(0.118)	(0.189)	(0.178)	(0.557)	(0.539)	(0.265)	(0.899)
Observations	524	524	524	487	487	487	685	685	685	173	173	173
R-squared	0.434	0.385	0.236	0.438	0.413	0.203	0.415	0.363	0.232	0.456	0.454	0.266
Number of ISINs	164	164	164	154	154	154	210	210	210	57	57	57
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Two-tailed p-values based on robust standard errors provided in parentheses.

Table 3. Regression results for continual CEP improvements and CFP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	ROA	TQ	ROA	TQ	ROA	TQ	ROA	TQ
ΔENERGY	0.0271	0.202						
	(0.000)	(0.000)						
ΔWATER			0.0291	0.194				
			(0.000)	(0.000)				
ΔCO_2					0.0282	0.191		
					(0.000)	(0.000)		
ΔWATERDISCHARGE							0.0299	0.208
							(0.000)	(0.000)
Size	-0.0164	-0.425	-0.009	-0.339	-0.0145	-0.370	-0.00652	-0.327
	(0.000)	(0.000)	(0.053)	(0.000)	(0.001)	(0.000)	(0.492)	(0.000)
Risk	-0.162	-1.001	-0.178	-1.128	-0.158	-0.867	-0.170	-0.731
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
R&D Intensity	-0.261	-0.733	-0.317	-1.955	-0.337	-0.924	-0.249	-1.109
	(0.000)	(0.129)	(0.000)	(0.010)	(0.000)	(0.095)	(0.028)	(0.083)
Observations	3,267	3,228	3,277	3,229	3,812	3,774	1,611	1,583
R-squared	0.164	0.234	0.186	0.239	0.179	0.228	0.167	0.230
Number of ISINs	660	649	631	624	751	750	354	349
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Two-tailed p-values based on robust standard errors provided in parentheses.

Table 4. Robustness checks on continual CEP improvements and Asset4 Environmental Pillar scores

VARIABLES	(1) Asset4	(2) Asset4	(3) Asset4	(4) Asset4
ΔENERGY	-1.082			
	(0.476)			
ΔWATER		1.141		
		(0.510)		
ΔCO_2			0.287	
			(0.861)	
ΔWATERDISCHARGE				-2.377
				(0.272)
ROA	-5.709	-17.62	-15.75	-1.290
	(0.617)	(0.169)	(0.174)	(0.941)
Size	6.160	7.782	6.483	8.761
	(0.000)	(0.000)	(0.000)	(0.008)
Risk	-7.095	-4.443	-3.556	-6.678
	(0.215)	(0.481)	(0.533)	(0.481)
R&D Intensity	-1.304	-1.990	6.095	16.919
	(0.907)	(0.872)	(0.607)	(0.351)
Observations	3,184	3,201	3,701	1,569
R-squared	0.1165	0.1233	0.1134	0.1556
Number of ISINs	655	629	750	352
Firm FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Two-tailed p-values based on robust standard errors provided in parentheses.

Table 5. Robustness checks on continual and cumulative CEP improvements and CFP

VARIABLES	(1) ROA	(2) TQ	(3) ROA	(4) TQ	(5) ROA	(6) TQ	(7) ROA	(8) TQ
ΔENERGY	0.031	0.136						
	(0.000)	(0.000)						
ΔENERGY_cumulative	0.002	0.045						
	(0.070)	(0.000)						
ΔWATER			0.035	0.115				
			(0.000)	(0.003)				
ΔWATER_cumulative			0.001	0.047				
			(0.401)	(0.000)				
ΔCO_2					0.026	0.090		
					(0.000)	(0.009)		
ΔCO_2_cumulative					0.003	0.038		
					(0.005)	(0.000)		
ΔWATERDISCHARGE							0.033	0.143
							(0.000)	(0.005)
ΔWATERDISCHARGE_cumulative							-0.001	0.021
							(0.532)	(0.203)
Size	-0.0220	-0.452	-0.012	-0.366	-0.018	-0.375	-0.0141	-0.373
	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.000)	(0.115)	(0.000)
Risk	-0.121	-0.631	-0.130	-0.688	-0.120	-0.578	-0.0946	-0.238
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.223)
R&D Intensity	-0.218	-0.325	-0.262	-1.011	-0.276	-0.392	-0.228	-0.330
	(0.000)	(0.391)	(0.001)	(0.063)	(0.000)	(0.290)	(0.045)	(0.462)
ROA (lagged)	0.264		0.263		0.263		0.336	
	(0.000)		(0.000)		(0.000)		(0.000)	
Tobin's Q (lagged)		0.407		0.404		0.404		0.422
		(0.000)		(0.000)		(0.000)		(0.000)
Observations	3,143	3,125	3,153	3,130	3,671	3,659	1,548	1,522
R-squared	0.235	0.408	0.250	0.418	0.244	0.404	0.258	0.384
Number of ISINs	645	639	618	617	734	739	344	341
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES

Two-tailed p-values based on robust standard errors provided in parentheses.