Article published in Journal of Industrial Ecology (2017)

https://doi.org/10.1111/jiec.12573

Economic Development Matters: A Meta-Regression Analysis on the Relation between Environmental Management and Financial Performance*

Markus Hang[a] (Corresponding Author)

Markus.Hang@mrm.uni-augsburg.de

Jerome Geyer-Klingeberg[a]

Jerome.Geyer-Klingeberg@mrm.uni-augsburg.de

Andreas W. Rathgeber[a]

Andreas.Rathgeber@mrm.uni-augsburg.de

Stefan Stöckl[b]

Stefan.Stoeckl@icn-groupe.fr

[a] *Author's affiliation*

Institute of Materials Resource Management
Faculty of Mathematics and Natural Sciences
University of Augsburg
Alter Postweg 101
86135 Augsburg, Germany

[b] *Author's affiliation*

Department of Finance, Audit, Accounting and Control
ICN Business School Nancy-Metz (Grande école) - CEREFIGE
3 place Edouard Branly
57070 Metz, France

Accepted version
Journal of Industrial Ecology, 2018, 22(4), pp. 720-744.

2017

* We express our gratitude to the Editor of the Journal of Industrial Ecology and four anonymous referees for their valuable feedback that helped to significantly improve the manuscript. Moreover, we gratefully thank Jan Endrikat, Thomas Guenther, Tom Stanley, Phu Nguyen-Van, as well as all other seminar participants at the 2015 Colloquium of the Meta-Analysis of Economics Research Network (MAER-Net), 78th Annual Meeting of the German Academic Association for Business Research (VHB), and 65th Annual Meeting of the French Economic Association (AFSE) for their helpful comments and suggestions. We acknowledge support from the Bavarian-French academic centre (Bayerisch-Französisches Hochschulzentrum/Centre de Coopération Universitaire Franco-Bavarois).

Abstract

Although the existing body of empirical literature on the relation between corporate environmental performance (CEP) and corporate financial performance (CFP) is continuously growing, results are still inconclusive about this fundamental question in industrial ecology. Comparisons are difficult because of various estimation methods as well as the overall heterogeneous and complex interaction between the two constructs, but especially because of country-specific data sets. Consequently, we raise the question of whether regional differences are the driving force buried underneath the inconclusiveness. Therefore, the aim of this article is to explore this heterogeneity by aggregating 893 existing results from 142 empirical primary studies that are based on more than 750,000 firm-year observations. Our findings suggest a convex impact of a country's economic development on the magnitude of the CEP-CFP effect (i.e., the effect is positive in developing countries, disappears in emerging countries, and is again positive in highly developed countries). We also find that the overall positive relation strengthens for market-based CFP measures and diminishes for countries with civil law systems, firms from the service sector, reactive environmental activities, and process-based CEP measures. Further, several aspects of the examined data sample and the inclusion of relevant control variables explain heterogeneity in previous research results.

Keywords: corporate environmental performance (CEP), corporate financial performance (CFP), heterogeneity, industrial ecology, meta-analysis, moderating effects

JEL Classifications: C33, Q51, Q53

1. Introduction

> *"We conclude, from our findings, that a unified theory cannot be derived because researchers always approach the topic from their limited perspective. Future research should test the synopsis by thoroughly assessing the large amount of empirical knowledge that already exists."*
>
> Guenther and Hoppe (2014, 701)

Over the past 40 years, hundreds of studies have found evidence for a positive relationship between corporate environmental performance (CEP) and corporate financial performance (CFP). However, empirical results largely vary due to differences in their variable measures, sample compositions, estimation methods, as well as the overall heterogeneous and complex interaction between CEP and CFP (Elsayed and Paton 2005; Guenther and Hoppe 2014; Guenther et al. 2012; McWilliams and Siegel 2000, Yang et al. 2011).

Numerous qualitative and quantitative literature reviews aim to aggregate the available empirical findings and to explain their variability across studies. The two works that are closest to this study are from Endrikat et al. (2014) and Horváthová (2010). Endrikat et al. (2014) focus on the correlation and causality of the CEP-CFP relation using weighted mean effects and subgroup analysis. Further, they specifically look at the impact of frequently used control variables in primary studies. Horváthová (2010) applies a vote counting approach to examine studies with different statistical approaches and finds the latter to explain the probability to obtain certain primary study results. So far, researchers mainly use univariate frameworks to analyze differences in measurement and methodology of the primary studies. This approach often neglects other important determinants of the CEP-CFP link such as the regional, industrial, and temporal conditions, as well as publication characteristics and sources of misspecification bias. Literature regards these variables as important factors for the interaction between CEP and CFP (Dixon-Fowler et al. 2013; Guenther et al. 2012; Guenther and Hoppe 2014; Molina-Azorín et al. 2009; Wagner et al. 2001). In addition, Wagner and Schaltegger (2003) as well as Vishwanathan (2010) underline that these variables interact with each other leading to a dependency structure among the moderating factors of the CEP-CFP relation. As a result, several authors recently raised the question of the multidimensional conditions that foster the prospects of success for corporate environmental activities (Delmas et al., 2011; Dixon-Fowler et al. 2013; McWilliams et al. 2006; Vishwanathan 2010). We address this issue by examining multiple determinants of the CEP-CFP nexus within a meta-regression analysis (MRA) that builds on a sample of 893 existing results drawn from 142 empirical primary studies. In that respect, we simultaneously test the impact of 42 explanatory variables capturing various aspects of a country's economic development and legal environment, measurement of the CEP and CFP variable, differences in the underlying data samples, estimation methods, and inclusion of important control variables. This analysis provides new insights into the drivers of heterogeneity in the empirical results on the CEP-CFP relation.

2. Related meta-analytical literature

Table 1 summarizes recent meta-analyses on the sources of the variability in the results for the CEP-CFP relation, where Endrikat et al. (2014) and Horváthová (2010) present the studies closest to our work. Obviously, differences exist among their findings as well as in their study designs. To elaborate on the contribution this paper makes, we discuss the existing meta-analyses in terms of differences in (1) the analyzed explanatory factors, (2) the chosen effect size, (3) their data sample, and (4) the applied meta-analytical approach.

Guenther and Hoppe (2014, 69) find that primary studies have "limited perspectives" due to their focus on partial aspects of the CEP-CFP relation. A similar picture can be found in terms of meta-analyses. Each study uses a specific selection of explanatory factors, whereby differences in methodology and the measurement of the CEP and CFP variables are the most frequently examined dimensions. However, literature suggests that also regional and macroeconomic conditions as well as publication and data characteristics might heavily affect the empirical results (Dixon-Fowler et al. 2013, Guenther et al. 2012; Guenther et al. 2014; Molina-Azorín et al. 2009; Wagner et al. 2001).

Second, the samples of included primary studies indicate remarkable differences. Existing meta-analyses built on between 20 and 274 primary studies. Stanley and Doucouliagos (2012) point out that a meta-study should aim at finding the complete population of primary studies in order to obtain meaningful results.[1] As the sample sizes used by Guenther et al. (2012) and Endrikat et al.

Table 1: Comparison with prior meta-analyses investigating the explanatory factors of the CEP-CFP relationship

Study name	Analyzed categories of explanatory variables	Number of studies	Date range	Grey literature[1]	Number of effect sizes	Numerical effect size	Method	Significant key findings for the CEP-CFP relation and analyzed moderating factors
Albertini 2013	• Regional differences • Industrial differences • Differences in variable operationalization • Data characteristics	52	1975-2011	No	205	Pearson correlation coefficient	Univariate meta-analysis, subgroup analysis	1) Overall positive relation between CEP and CFP 2) Differences in variable operationalization, regional differences, and duration determine the relation: • Rest of the world > US/Canada > EU • Accounting based CFP > market-based CFP • Non-longitudinal studies > longitudinal studies
Dixon-Fowler et al. 2013	• Industrial differences • Differences in variable operationalization • Data characteristics • Differences in estimation methods	39	1970-2009	No	202	Pearson correlation coefficient	Univariate meta-analysis, subgroup analysis	1) Overall positive relation between CEP and CFP 2) Sample characteristics and measurement characteristics matter: • Small firms > large firms • US > rest of the world • Market-based CFP > accounting-based CFP
Endrikat et al. 2014	• Differences in variable operationalization • Data characteristics • Differences in estimation methods • Control variables	149/117[2]	1970-2012	Yes	245/208[2]	Pearson correlation coefficient/partial correlation coefficient[2]	Univariate meta-analysis, subgroup analysis	1) Overall positive relation between CEP and CFP 2) Measurement characteristics, control variables, endogeneity, sample characteristics, and the observation period determine the effect: • Proactive environmental activities > reactive environmental activities • Financial risk not controlled > financial risk controlled • Endogeneity not addressed > endogeneity addressed • Single industry sample > cross sectional sample
Guenther et al. 2012	• Differences in variable operationalization	274	1970-2010	No	321	Categorical variable	Narrative review, vote counting	1) Overall positive relation between CEP and CFP 2) Measurement characteristics matter: • Strategic CEP measures/ratings > environmental reporting/events/operational CEP • Questionnaire based CFP > accounting based CFP/market based CFP
Horváthová 2010	• Differences in variable operationalization • Data characteristics • Differences in estimation methods	37	1978-2007	Yes	64	Categorical variable	Vote counting, ordered probit model	1) Overall positive relation between CEP and CFP 2) Measurement, estimation, and sample characteristics partially matter: • regression analysis/panel data analysis > correlation coefficient/portfolio studies • Longitudinal studies > non-longitudinal studies • US/Canada > rest of the world • Process-based CEP > outcome-based CEP • Common law countries > civil law countries
Murphy 2002	• Differences in variable operationalization	20	1992-2002	No	20	-	Narrative review	1) Overall positive relation between CEP and CFP 2) Measurement characteristics matter: • Environmental disclosure ≠ CEP • Proactive CEP positively related to CFP

This study	• Regional differences • Industrial differences • Differences in variable operationalization • Data characteristics • Publication characteristics • Differences in estimation methods • Inclusion of important control variables	142	1970-2015	Yes	893	Partial correlation coefficient	Multiple random effects MRA with cluster robust errors	1) Overall positive relation between CEP and CFP 2) Regional differences, industrial differences, and differences in variable operationalization[3]: • Developing countries[4] > rest of the world • BRICS[5] < rest of the world • Asia < rest of the world • US > rest of the world • Decreasing effect size with increasing GDP per capita • Decreasing effect size with increasing GDP per capita growth rate • Common law countries > civil law countries • Industrial firms > service industry • Outcome-based CEP measure > process-based CEP measure • Market-based CFP measure > accounting-based CFP measure • Proactive environmental activities > reactive environmental activities Further variables: • Firm size not controlled > Firm size controlled • Capital intensity not controlled > capital intensity controlled • Multinational studies > single country studies • Pre 2000 > post 2000 • 2000-2009 > remaining time period[5] • Increasing effect size with growing number of authors

This table sums up prior research on the moderating factors of the CEP-CFP relation and compares them with the study at hand. The second column contains the categories of explanatory variables which correspond to the analyzed factors. Further, the underlying number of primary studies is given in column 3. Column 4 comprises the corresponding date range from which the study sample is drawn, and column 5 indicates if the study explicitly incorporates grey literature. Column 6 encompasses the number of extracted effect sizes across all studies. Finally, the last column sums up the results concerning the (1) CEP-CFP relation, as well as (2) concerning the moderating factors that were analyzed in the respective study. Therein, ">" denotes that the CEP-CFP effect is larger for the former type of effect size. "≠" denotes that the review identified a different behavior for the two types of effect sizes.

[1] Following Nelson and Kennedy (2009, 349), we define "grey literature" as unpublished studies like working papers, conference proceedings, or doctoral theses.
[2] The first number of studies/effect sizes corresponds to the analysis based on Pearson correlation coefficients; the second refers to the meta-analysis using partial correlations.

[3] These results have to be interpreted in combination with the other variables. See also the example in the discussion section.

[4] Data for this classification is taken from the Organization for Economic Co-operation and Development (OECD): https://www.oecd.org/dac/stats/documentupload/DAC List of ODA Recipients 2014 final.pdf.

[5] The BRICS cover the following countries: Brazil, China, India, Russia, and South Africa.

(2014) show, the body of empirical evidence on the CEP and CFP relationship is huge. In contrast, some of the other previous reviews build on comparably smaller samples. In addition, more than half of the studies completely ignore grey literature. This exclusion of unpublished studies bears a number of risks: As such, it might result in insufficient variation in the explanatory variables (Stanley et al. 2008), which is indeed necessary to explore the sources of heterogeneity in a multiple regression model. Moreover, unpublished work often covers recent data. Thus, their exclusion reduces the observed time span and the possibility to identify temporal effects.

Third, the preferred effect size in existing meta-analyses is the Pearson correlation coefficient (Albertini 2013; Dixon-Fowler et al. 2013; Endrikat et al. 2014). However, using zero-order correlations might lead to a false picture of the underlying effect as it might contain spurious interactions due to the multidimensional behavior of the analyzed factors (Stanley and Doucouliagos 2012). Horváthová (2010) highlights this bias in the case of the CEP-CFP measurement by comparing the results of different effect sizes. This implies that the applied effect size should only measure the marginal or partial effect of the relation between CEP and CFP.

Finally, the applied methodologies cover qualitative literature reviews, vote counting procedures, univariate meta-analyses, and univariate subgroup analyses. According to Borenstein et al. (2009) and Hedges and Olkin (1985), qualitative reviews are affected by subjective judgements of the authors (e.g., in the selection of the study sample or the weighting of the studies). To integrate results from various estimation models and analyze the respective differences, Horváthová (2010) applies a vote counting procedure in the special form of an ordered probit model.[2] Here, vote counting results serve as categorical dependent variable which is defined as 1 for a positive and statistically significant relationship between CEP and CFP, 0 for an significant relation, and −1 for a significantly negative relationship. Such an approach leads to a loss of information about the magnitude and precision of the effects as it concentrates on the analysis of the frequency of certain study outcomes (Stanley and Doucouliagos 2012). Finally, univariate approaches and subgroup analyses as applied by Endrikat et al. (2014) ignore the multidimensional dependency structure among the explanatory variables (Bender et al. 2008), which is indeed regarded to be important (Wagner and Schaltegger 2003; Vishwanathan 2010).

The goal of this study is to close the mentioned gaps by investigating a broad set of explanatory variables in a multidimensional manner using MRA methodology, including a large sample of primary studies capturing also grey literature, and using an effect size that measures the partial relation between CEP and CFP while holding other factors constant.

3. Theory of the CEP-CFP relation and its moderating factors

Although recent reviews show that definitions of CEP and CFP widely vary across empirical literature (among others, Albertini 2013; Endrikat et al. 2014; Guenther et al. 2012), there is a generally uniform understanding of the two artifacts. CEP[3] is understood as "the absolute performance of a firm with regard to the environment, i.e. its environmental impact" and CFP as "the result of a company's activities regarding its targets: liquidity, profit and strategic profit potential" (Guenther et al. 2012, 280). In contrast to this uniformity, three causal directions are plausible for the relation between CEP and CFP. CEP influences CFP, CFP affects CEP, and a bidirectional link. Literature is also inconsistent about the sign of the relationship, which might be negative, neutral, or positive. Table 2 summarizes a selection of the main arguments used in empirical literature, where we generally follow the structure from Preston and O'Bannon (1997) as well as Waddock and Graves (1997).[4]

Looking at CEP influencing CFP, representatives of the tradeoff hypothesis argue that environmental management causes an additional cost burden for firms which does not exceed the related benefits (Levitt 1958). Consequently, there is only a choice between investing either in a profitable, or in a responsible firm (Bragdon and Marlin 1972). Furthermore, McWilliams and Siegel (2001) develop the supply and demand model that assumes that CFP is independent from CFP, and vice versa. Hence, an optimal investment in environmental issues can be determined by usual cost-benefit analyses. This synthesis of interests considers not only the maximization of profitability demanded by shareholders, but also the claims of stakeholders for environmental responsibility such as those of customers or communities (McWilliams and Siegel 2001). Next, Porter and van der Linde (1995a, 1995b) argue that properly designed legal restrictions promote innovations that compensate the additional costs of regulation. The natural resource-based view

(NRBV) supports this theory by suggesting competitive advantages through new resources and capabilities from environmental responsibility (Barney 1991; Wernerfelt 1984). This also includes long-term profit maximization (Davis 1973) and the consideration of stakeholder interests (Donaldson and Preston 1995).

Table 2: Typology of theoretical CEP-CFP relations

<table>
<tr><th rowspan="2">Causal sequence</th><th colspan="3">Direction</th></tr>
<tr><th>negative</th><th>neutral</th><th>positive</th></tr>
<tr><td>CEP → CFP</td><td>Tradeoff hypothesis
Financial returns from environmental engagement do not completely compensate the investments by the firm (Levitt 1958; Bragdon and Marlin 1972).</td><td rowspan="2">Supply and demand model
An optimal investment in environmental issues can be determined by cost-benefit analysis (Barnett 2007; McWilliams and Siegel 2001).</td><td>Porter hypothesis/Natural resource-based view/ Instrumental stakeholder theory/Social impact hypothesis
Firms profit from environmental activities through competitive advantages, enhanced efficiency, or meeting stakeholder expectations (Barney 1991; Donaldson and Preston 1995; Esty and Porter 1998; Hart 1995; Latané 1981; Orlitzky et al. 2003; Porter and van der Linde 1995a, 1995b; Wernerfelt 1984).</td></tr>
<tr><td>CEP ← CFP</td><td>Managerial opportunism hypothesis
Managers follow their own private targets in their daily business, which may not be in line with shareholder interests (Posner and Schmidt 1992; Preston and O'Bannon 1997; Weidenbaum and Vogt 1987).</td><td>Slack resources theory
The availability of slack resources via good financial performance enables firms to invest in environmental programs (Bourgeois 1981; Kraft and Hage 1990).</td></tr>
<tr><td>CEP ↔ CFP</td><td colspan="3">Virtuous circle
Firm enhances its intangible assets by investing in environmental programs. This leads to financial benefits that must reinvested in intangible assets for a further improvement of CEP (Surroca et al. 2010; Waddock and Graves 1997).</td></tr>
</table>

This table classifies the theories concerning the direction of the effect and the supposed causal sequence between CEP and CFP.

For the opposite causality, the managerial opportunism hypothesis proposes a negative impact of CFP on CEP (Weidenbaum and Vogt 1987). Hence, when a firm performs well, managers follow their own targets by increasing their income through reductions of environmental investments. On the other hand, managers want to compensate their disappointment or even justify bad corporate performance by expanding a firm's environmental expenditure (Preston and O'Bannon 1997). In contrast to these assumptions, representatives of the slack resources theory suggest a positive impact of CFP on CEP. As such, the availability of slack resources via good financial performance enables firms to invest in environmental programs, which enhances internal resources and creates new capabilities as well as comparative advantages (Kraft and Hage 1990).

Waddock and Graves (1997) then reconciled these two views by drawing on the concept of the "virtuous circle", which assumes a simultaneous relation between CEP and CFP (Waddock and Graves 1997). Accordingly, Surroca et al. (2010) argue that a firm enhances its intangible assets by investing in environmental programs. This leads to financial benefits that must be reinvested in intangible assets for a further improvement of CEP.

The continuous development of new theories and the theoretical inconsistency might be a reason for the heterogeneous picture provided by the empirical literature. Guenther and Hoppe (2014) point out that the varying behavior of the CEP-CFP relation might also exist due to the multidimensional dependency structure including further moderating, mediating, or confounding factors, which we analyze in this study.

Regional differences

Kerret and Shartzvald (2012), Sotorrio and Sánchez (2008), as well as Yang et al. (2011) state that regional differences influence the effectiveness and profitability of environmental practices. These include economic, legal, social, and political factors (Dixon-Fowler et al. 2013).

Di Vita (2009) as well as Zhu et al. (2007) put forward that a country's economic conditions determine environmental responsibility and activities. For example, the environmental Kuznets curve describes a U-shaped relation between the GDP per capita and the intensity of environmental emissions (Dinda 2004). Following this proposition, the CEP-CFP relation should be weaker in emerging countries as opposed to developing countries, and increase again with growing wealth. Thus, discrepancies might originate from different geographical zones and economic situations (Bernauer and Boehmelt 2013). In a similar vein, Sotorrio and Sánchez (2008) reveal that in European countries, the aim of achieving better financial performance through increased environmental responsibility is much higher than in the US. Moreover, Yang et al. (2011) state that CEP has a significant positive impact on CFP in developed countries.

Another reason might arise from different legislative actions (Allouche and Laroche 2007; Brouwers et al. 2014) which influence a firm's attitude against pollution prevention (McGrath and Gilbert-Miller 1999). Further, different law systems may also affect the environmental performance due to varying shareholder protection. By encouraging financial investments, effective stakeholder protection in common law countries increases the growth of the income per capita and lowers the cost of environmental activities through reduced interest rates and discount rates (Di Vita 2009).

Hypothesis 1: Regional differences explain variation in the reported estimates for the CEP-CFP relation.

Industrial differences

Shetty (1979) observes that several researchers found an association between the level of environmental responsibility and industry-level differences. As the analysis of multi-industry samples may obscure such effects (Guenther and Hoppe 2014), one strand of the literature explicitly refers to single industries. Here, the manufacturing industry (among others, Iwata and Okada 2011; King and Lenox 2001, 2002; Kock et al. 2012; Ragothaman and Carr 2008; Tatsuo 2010; Wagner 2005), and companies form the service industry (among others, Inoue and Lee 2010; Kassinis and Soteriou 2005; Llach et al. 2013) are widely represented. As stated by Villiers et al. (2011), firms in "dirty (sensitive) industries" like the manufacturing sector can more effectively deal with environmental pollution, which might result in a stronger CEP-CFP relation. Yang et al. (2011) show that there might be discrepancies in the handling of environmental pollution between industries comprising small and medium-sized enterprises compared to those comprising large firms. The latter may have more resources favoring research and development and new technologies (Dixon-Fowler et al. 2013).

Hypothesis 2: Industrial differences explain variation in the reported estimates for the CEP-CFP relation.

Differences in variable operationalization

Various studies (among others, Busch and Hoffmann 2011; Endrikat et al. 2014; Griffin and Mahon 1997; Ilinitch, et al. 1998) suspect differences in the variable definitions to significantly influence empirical findings. Especially for CEP there exists a broad range of proxy variables. These encompass, among others, the extent of pollution prevention (Berrone and Gomez-Mejia 2009), the amount of toxic releases or greenhouse gas emissions (Day et al. 1997; Delmas and Nairn-Birch, 2011), the certifications of implemented environmental management systems (Heras-Saizarbitoria et al. 2011), or environmental ratings and rankings (Inoue and Lee 2010). To reach a reasonable taxonomy in the prevailing complexity of CEP measurement, the proxy variables in this paper are categorized by their strategic level and quantifiability, which is in line with management literature (among others, Ginsberg, 1988; Miller & Friesen, 1983; Trumpp, Endrikat, Zopf, & Guenther, 2015). Therein, "process-based" measures refer to CEP on a management or process level. These include management practices, environmental policies, or environmental innovation. On the other hand, "outcome-based" measures try to quantify the real impacts of these efforts by measuring the amount of emissions, the ratio of recycled waste to total waste, or energy consumption. Among others, Busch and Hoffmann (2011) provide evidence for different effects of process-based and outcome-based measures on CFP.

Moreover, the starting point of corporate environmental activities is clustered in either "proactive" or "reactive" efforts (Hart 1995; Henriques and Sadorsky 1999; Walls et al. 2011). Proactive solutions aim to prevent environmental pollution through building on green process design, special capabilities, and resource combinations of the firm (Walls et al. 2011). Firms might especially profit from reduced risk associated with higher energy consumption and increased revenues through material savings or waste avoidance (Sen et al. 2015). On the contrary, firms only use "end-of-pipe solutions" like air filters or water clearers to comply with regulations and laws in order to minimize costs, risks, and liabilities (Yamaguchi 2008, Roome 1992). Walls et al. (2011) also reveal the differing nature of the interaction of proactive and reactive environmental activities with CFP.

Furthermore, literature generally divides CFP into "market-based" (for example, Tobin's Q, stock return, or market value of a firm) and "accounting-based" (for example, return on assets, return on equity, or return on sales) measures for the financial situation of a firm (Guenther et al. 2012). This has consequences on the direction of the CEP-CFP relation as pronounced by Delmas and Nairn-Birch (2011). The distinctive characteristics cover the forward-looking properties of market-based measures, taking intangible and future values of environmental efforts into account, as well as the backward-looking properties of accounting-based measures, which consider tangible costs and revenues induced by environmental activities.

Hypothesis 3: Differences in variable operationalization explain variation in the reported estimates for the CEP-CFP relation.

Further explanatory variables

A further group of moderating factors corresponds to publication characteristics which are typically included in MRAs in order to control for study quality not captured by the other explanatory variables (Doucouliagos and Ulubasoglu 2008; Havránek and Irsova 2011; Stanley et al. 2013). We incorporate the number of authors (for previous application of this variable, see Ankamah-Yeboah and Rehdanz 2014; Garcia-Quevedo 2004; Gurdgiev et al. 2010, among others) as well as the number of citations (see for previous application of this variable, among others, Demena and van Bergeijk 2016; Havránek et al. 2015; Kysucky and Norden 2016) in the analysis, as we can objectively derive them from each study. For the number of authors, we assume that a larger group could potentially increase the quality of the results due to more persons checking the validity of the outcomes (Tucker et al. 2016). The number of citations is preferred over the journal impact factor, as it considers study-specific quality characteristics and it is available for each study, including unpublished works.

Additionally, we further take data characteristics into accounts. A common variable is the mean year of data observed in the primary studies (see, among others, Goerg and Strobl 2001; Stanley and Jarrel 1998). We integrate this variable in our analysis as we assume that the CEP-CFP relation changed over time due to the development in corporate expenditures for environmental activities, governmental subsidies, as well as customers' demand for environmentally friendly producers. Following recent MRAs by Doucouliagos and Ulubasoglu (2008) and Valickova et al. (2015), we include the number of firms, the length of the observation period, the number of countries, and the number of industries. The sample size and the length of the observation period are measures for the precision of a certain finding and the distinction between short- and long-term effects (Inoue and Lee 2011; Stanley et al. 2013). As pointed out by Griffin and Mahon (1997), the differences in the CEP-CFP nexus gradually blur as the number of simultaneously analyzed countries and industries increases. We account for this effect by including variables for the number of examined countries and industries.

Moreover, estimation differences among primary studies could fundamentally determine the deviations of the findings. Simple regression techniques tend to bias the particular effect

(Horváthová 2010), as for example heterogeneity or dependencies are ignored. We incorporate variables to control for the potential application of simple regression approaches as well as the consideration of fixed and random effects. Furthermore, as shown in Table 2, the endogeneity problems between CEP and CFP constitute a critical aspect. Primary studies often address this issue by using lagged variables (Cohen et al. 1997; Konar and Cohen 2001). Hence, we code a variable for the number of lagged years between CEP and CFP. Moreover, we add a further dummy variable that controls for the explicit consideration of endogeneity between CEP and CFP by, for example, lagged variables or instrumental variables estimations.

A special capability of MRA is the correction of the initial estimates for model specification errors arising from the omission of important control variables (Stanley and Doucouliagos 2012). Thus, we integrate the relevant control variables used in primary studies in the MRA model as dummy variables. In detail, the MRA model accounts for the inclusion of a variable for research and development expenditure, advertising intensity, capital intensity, financial risk, sales growth, firm size, and industry dummy variables.

Hypothesis 4: (*a*) Data characteristics, (*b*) publication characteristics, (*c*) estimation differences, and (*d*) model misspecifications explain variation in the reported estimates for the CEP-CFP relation.

4. Empirical analysis

The Meta-Analysis of Economics Research Network (MAER-Net) provides established guidelines for the reporting of meta-regression results (Stanley et al. 2013). As described in Appendix B, this study complies with all compulsory aspects mentioned in these guidelines. We solve the controversial point concerning the coding of the sample studies by developing a common understanding of all variables to be coded. Afterwards one author coded the full sample of studies. During the data preparation process every uncertainty about a variable has been discussed among the authors until a common consensus for the discordant data point has been found. Further information regarding the coding of the explanatory variables is available from Appendix C.

Data search and preparation

For the construction of the meta-data set, we applied the following search strategy: We started with the set of 149 primary studies reported in Endrikat et al. (2014). As the latest study in their sample is from 2012, we replicated the search strategy with minor changes[5] for the period from 2012 till the end of 2015. We searched a variety of major electronic databases for published (ABI/Inform Complete, Business Source Complete, EconLit, GreenFILE, and ScienceDirect) and grey literature (Social Science Research Network) using a broad search term consisting of 24 different keywords[6]. This extended search yielded another 49 relevant studies. More details on the data search for the extended period can be found in Appendix D.

In order to produce a homogeneous sample of primary studies over the whole period from 1970 to 2015, we filtered the list of relevant studies against specific criteria a study had to satisfy to be included in the final sample. The applied inclusion and exclusion criteria can be categorized into requirements concerning (1) effect size measurement, (2) methodology, and (3) data. The first category defines that a study has to measure the relation between CEP and CFP using firm-level data[7]. Studies using disclosure data as a proxy for CEP are excluded from the sample. This is due to the fact that contemporary empirical analyses reveal an inconsistent and even contradictory behavior in contrast to other common measures like emission intensity (among others, Aragón-Correa et al. 2015; Hughes et al. 2001; Patten 2002). For the measurement of CFP, we do not consider perceptual measures as they are not equivalent to other objective measures (Richard et al. 2009).[8] The second category combines the requirements for an empirical and multidimensional analysis of the relation between CEP and CFP. As the majority of effect sizes in the CEP-CFP literature are continuous measures, event studies and results from probit/logit models are excluded from the analysis, since a meaningful aggregation across many primary studies requires the effect sizes to be comparable (Stanley and Doucouliagos 2012). The third category sums up the necessary data for the calculation of effect sizes. In detail, regression coefficients, the corresponding standard error and sample sizes have to be given or at least be reproducible from the reported data.

Applying these criteria to the list of 198 relevant publications then produces our final sample of 142 primary studies. A full list of all included studies can be found in Appendix E. Further information about the sample is provided in Appendix F, G, and H.[9] The study sample has also been checked for independency. Therefore, we delete studies from the same (co-)authors with identical samples to avoid double counting (Stanley and Doucouliagos 2012). In contrast, overlapping or equal data sets in studies from different authors or studies from the same authors using different data sets are defined as independent (Doucouliagos and Ulubasoglu 2008; Hunter and Schmidt 2004).

As effect size, we use the partial correlation coefficient (*r*), which is directly drawn from regression results reported in primary studies.[10] It measures the intensity of the CEP-CFP relation, while holding all other factors from the regression model constant.[11]This effect size contains several advantages. First, compared to the zero-order correlations as frequently used in management research, the effect of interest is corrected for spurious influences and thus allows a detailed analysis of misspecification bias by investigating differences in estimation models and integrated control variables. Second, the use of the zero-order correlations would end up in a smaller set of studies that could be included, which might represent a biased sample of the whole research record. Moreover, the use of partial correlations allows to integrate multiple estimates per study that are observable as most studies report results for several regression models (e.g., for different specifications or subsamples), which is usually not the case for zero-order correlations. This increases the sample size, enables the analysis of within-study variation, and enhances the variability in the data set (Chaikumbung et al. 2016). Finally, the broad sample allows the simultaneous integration of the relevant explanatory factors, which is especially vital for the analysis of the CEP-CFP relation (Endrikat et al. 2014; Horváthová 2010).

Nevertheless, the partial correlations have drawbacks to be considered and evaluated. To account for the fact that the partial correlation coefficients are not normally distributed, we additionally calculate the results based on Fisher's *z*-transformed values as a robustness test and transform them back to partial correlation coefficients for reporting. As the results are fairly the same as for the non-transformed effect size, we report them in Appendix M. For a comparison of our results with previous reviews, we include a robustness test using the same categorical variable as proposed by Horváthová (2010).

As an alternative to the partial correlations, Aloe and Becker (2012) recently proposed the use of semi-partial correlations (r_{SP}) as effect size due to its superior characteristics when sampling estimates from models with multiple regressors. However, to the best of our knowledge, this effect size has not been used in other MRA applications, which is probably driven by several shortcomings regarding data availability. Compared to the partial correlation, r_{SP} requires the value of the R^2 statistic as input which is indeed not reported in all studies and reduces the total number of effect sizes by 21%. Further, the R^2 statistic is often available from simple regression models. Thus, newer studies and advanced econometric methods are underrepresented leading to a biased sample of primary studies. Thus, we rely on the partial correlations as effect sizes in our baseline models, but report the results for the semi-partial correlations as robustness test.

After checking the appropriateness of each regression analyses in the collected primary studies, the obtained 142 primary studies provide a total database of 893 effect size estimates, based on 241,104 firm observations and 757,154 firm-year observations respectively. Within this sample, 96 effect sizes are drawn from 12 unpublished studies. After the collection and preparation of the effect sizes, the explanatory variables have been coded. Their definition and the corresponding summary statistics are described in Table 3.

Table 3: List of explanatory variables

Variable code	Description	Mean	Std. dev.
Regional differences (*H1*)			
DVLP	= 1 if the study investigates only data from developing countries, 0 otherwise[1]	0.069	0.254
G8	= 1 if the study investigates only data from countries that are members of the Group of Eight (G8), 0 otherwise[2]	0.720	0.449
BRICS	= 1 if the study investigates only data from BRICS countries, 0 otherwise	0.056	0.230
ASIA	= 1 if the study investigates only data from Asia, 0 otherwise	0.059	0.236
EU	= 1 if the study investigates only data from the EU, 0 otherwise	0.199	0.400
AFRICA	= 1 if the study investigates only data from Africa, 0 otherwise	0.010	0.100
US	= 1 if the study investigates only data from the US, 0 otherwise	0.340	0.474
GDP[3]	Natural logarithm of country's average real GDP per capita over the observation period[4]	9.956	0.725
GDPGROW[3]	Geometric average of the growth rate of country's real GDP over the observation period[4]	0.056	0.051
COMMON	= 1 if a country's law system is classified as common law, 0 otherwise[5]	0.502	0.500
CIVIL	= 1 if a country's law system is classified as civil law, 0 otherwise[5]	0.348	0.477
Industrial differences (*H2*)			
SMALL	= 1 if the study investigates only data from small firms, 0 otherwise[6]	0.233	0.423
MANUF	= 1 if the study investigates data from the two-digit SIC codes 20-39, 0 otherwise	0.389	0.488
SERV	= 1 if the study investigates data from the two-digit SIC codes 60-99, 0 otherwise	0.104	0.306
DIRT	= 1 if the study investigates only data from dirty industries (two-digit SIC codes 10-49), 0 otherwise	0.502	0.500
Differences in variable operationalization (*H3*)			
PROCESS-BASED	= 1 if CEP is a process-based performance measure, 0 for outcome-based performance measures	0.432	0.496
MARKET-BASED	= 1 if CFP is a market-based performance measure, 0 for accounting-based performance measures	0.308	0.462

PROACT	= 1 if CEP measures proactive corporate activities, 0 otherwise	0.408	0.492
REACT	= 1 if CEP measures reactive corporate activities, 0 otherwise	0.068	0.252
CAUSAL	Counts the lagged years between the measurement of CFP and CEP in the regression analysis by subtracting the time period of the CEP variable from the time period of the CFP variable	0.319	0.835
Data characteristics (*H4a*)			
MILLENIUM	= 1 if the mean year of the firm data in a primary study is 2000 or later	0.643	0.479
DATERANGE	Measures the date range of primary data (in years)	4.361	3.314
NOFIRMS	Counts the number in the data sample	282.047	393.765
1980s	= 1 if the mean year of firm data is between 1980 and 1989, 0 otherwise	0.082	0.274
1990s	= 1 if the mean year of firm data is between 1990 and 1999, 0 otherwise	0.261	0.439
2000s	= 1 if the mean year of firm data is between 2000 and 2009, 0 otherwise	0.585	0.493
MLTNAT	= 1 if the study employs data from more than one country, 0 otherwise	0.143	0.351
NOCOUNT	Counts the number of countries covered by the primary data	2.207	4.461
MLTSECT	= 1 if the study employs data from more than one industry sector, 0 otherwise	0.487	0.500
NOSECT	Counts the number of industry sectors covered by the primary data	4.241	3.714
Publication characteristics (*H4b*)			
NOAUTHORS	Counts the number of authors	2.262	1.061
CITE	Counts the number of citations on Google Scholar as of June 02, 2015	155.046	321.729
Estimation differences (*H4c*)			
SMPL	= 1 if the study applies a simple regression approach, 0 otherwise	0.321	0.467
HTRG	= 1 if the study controls for unobserved heterogeneity in the statistical approach, 0 otherwise	0.197	0.398
ENDO	= 1 if the study controls for endogeneity between CEP and CFP by lagged variables or an appropriate statistical technique (like instrumental variable approach), 0 otherwise	0.429	0.495
Model misspecification (*H4d*)			
RD	= 1 if the study controls for intensity of research and development in the regression analysis, 0 otherwise	0.306	0.461
AD	= 1 if the study controls for advertising intensity in the regression analysis, 0 otherwise	0.171	0.377
CAP	= 1 if the study controls for capital intensity in the regression analysis, 0 otherwise	0.249	0.432
RISK	= 1 if the study controls for financial risk in the regression analysis, 0 otherwise	0.462	0.499
GROW	= 1 if the study controls for sales growth in the regression analysis, 0 otherwise	0.231	0.422
SIZE	= 1 if the study controls for firm size in the regression analysis, 0 otherwise	0.766	0.424
IND	= 1 if the study controls for industry effects in the regression analysis, 0 otherwise	0.296	0.457

This table presents the explanatory variables employed in the MRA. Beside the abbreviation of the variable name, a short description is given in the second column. The last two columns show the mean and the sample standard deviation for the particular variable.

[1] Data for this classification is taken from the Organization for Economic Co-operation and Development (OECD): https://www.oecd.org/dac/stats/documentupload/DAC List of ODA Recipients 2014 final.pdf.
[2] The G8 include Canada, France, Germany, Italy, Japan, Russia, United Kingdom, and United States.
[3] For the construction of the variable, we use the average value corresponding to the sample employed in the estimation of the regression coefficients.
[4] Data on GDP per capita and GDP growth is taken from the World Bank database: http://databank.worldbank.org.
[5] Data for the classification into common and civil law countries is taken from the World Factbook of the Central Intelligence Agency: https://www.cia.gov/library/publications/the-world-factbook/fields/2100.html.
[6] The data are judged based on firm size. If the market capitalization is smaller than one billion dollars, which is approximately the lower limit of the S&P 500, or has less than 1,000 employees, a firm is classified as small.

Methodology of meta-regression analysis

Meta-analysis often refers to the Hedges-Olkin-type meta-analysis (HOMA) approach applied, for example, in Endrikat et al. (2014). The HOMA is a univariate analysis that calculates weighted average effects of estimates observed from a sample of empirical studies on the same phenomenon and compares the mean effects among different subgroups. In contrast, the MRA approach applied in this article allows us to simultaneously test the impact of different explanatory variables in a multiple regression approach and thus to uncover which explanatory variables (sometimes referred to as moderator variables) cause the heterogeneity among the observed estimates (Stanley and

Jarrell 1989).[12] We use the isolated effect of interest (CEP-CFP relation) in the form of partial correlations r_{ij} as dependent variable and regresses them against a set of variables that are likely to explain the variation among the primary studies. The general MRA model can be described as follows:

$$r_{ij} = \beta_0 + \beta_1 SE(r_{ij}) + \sum_{k=1}^{K} \gamma_k Z_{ijk} + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0; SE(r_{ij})^2), \tag{1}$$

where $SE(r_{ij})$ is the standard error of the partial correlation r_{ij} of the ith estimate taken from the jth study, Z_{ijk} denotes the set of k explanatory variables defined in Table 3, and ε_{ij} is the error term.

One special capability of this model it to evaluate and correct for publication selection and misspecification bias that are both routinely found in empirical research (see, among others, Doucouliagos and Laroche 2009; Havránek and Irsova 2011; Stanley 2004). Publication selection arises when researchers favor results based on their statistical significance or because they are consistent with the majority of literature (Card and Krueger 1995). This biasing effect may distort statistical inferences and especially averaged effects from meta-analysis. Estimating the model without the Z vector provides the so-called "funnel asymmetry test" (FAT), which is used to test the presence and extent of publication selection bias (Stanley 2008).[13] Egger et al. (1997) suggest that the rejection of the null hypothesis, $\beta_1 = 0$, serves as a test for the presence of publication selection bias. This test also holds for the extended MRA model including the Z vector. Moreover, the intercept β_0 captures the size of the mean effect of CEP on CEP corrected for a potential publication bias. However, the mean effect must be interpreted conditional on the values for the Z vector. Second, misspecification of econometric models might arise from omitted variables, inappropriate estimation methods, or data limitations of primary studies (Stanley 2001). We investigate the existence of misspecification by integrating specific explanatory variables for differences in the estimation models and the inclusion of important control variables (see also Table 3).

For the estimation of the MRA model in Equation (1), we additionally account for the following aspects and report several alternative model specifications as robustness tests. (1) As the standard errors of the reported estimates usually have great variation across the primary studies, all estimations are carried out through weighted least squares (WLS) using the inverse of the squared standard errors as weights $1/[\mathrm{SE}(r_{ij})^2]$.[14] The WLS approach is commonly used and in line with recent research (Nelson and Kennedy 2009; Stanley and Doucouliagos 2015). It also implies that studies reporting lower standard errors and thus more precise results are given a larger weight in the MRA estimation (Hedges and Olkin 1985). (2) As we collect multiple estimates per study, we have to account for within-study dependencies among the effect sizes taken from the same study. Hence, we calculate robust standard errors clustered at the study level (Hedges et al. 2010). An alternative to the sampling of multiple estimates per study is given by the average set of estimates reported within the same study (Stanley and Doucouliagos 2012). Although this approach overcomes the dependency problem, it leads to a significant loss of information about the within-study variation that is indeed valuable to explain the heterogeneity in the existing literature. Furthermore, several additional explanatory factors could not be integrated in the analysis as they are not unambiguously assignable to the average effect size. Due to the disadvantage of average effect sizes, recent meta-analyses recommend the coding of multiple estimates per study together with an appropriate routine to account for dependencies in the estimation method (Havránek and Irsova 2011; Kysucky and Norden 2016; Valickova et al. 2015). We follow this best practice, but also report the results for the average-set of estimates as a robustness test. (3) We employ general-to-specific modeling to simplify the multiple MRA model. This is the common procedure in MRA research to obtain parsimonious models containing only the most important variables (see, for example, Abdullah et al. 2015; Doucouliagos and Stanley 2009; Valickova et al. 2015). After each re-estimation, the most insignificant variable is dropped from the model until the remaining t-statistics are all larger than one. (4) A random effects model is used to estimate the residual heterogeneity not covered by the explanatory variables, for example from variations of the effects due to unobserved heterogeneity on the firm-level (like management quality). The random effects weights are $1/[\mathrm{SE}(r_{ij})^2 + \tau^2]$, where τ is an estimate for the between-study variance. The random effects MRA is a frequently used model in existing research (see, for example, Chaikumbung et al. 2016; Havránek 2010; Stanley and Doucouliagos 2012).

5. Presentation of meta-analytical results

Test of excess heterogeneity, univariate meta-analysis, and publication selection bias

We conduct Cochran's Q test with the null hypothesis of homogenous effect sizes to test if there is heterogeneity in the observed effect size estimates. The test statistic Q_{total} of 4852.160, which is significant at the 1% level, clearly rejects the null hypothesis of homogeneity. This confirms that there is a significant proportion of excess variation beyond sampling error. Looking at the magnitudes of the significant values of $Q_{between}$ (2433.289) and Q_{within} (2418.871) reveals that heterogeneity is almost equally split between variation among studies ($Q_{between}$) and variation within studies (Q_{within}). This result encourages the integration of variables that might explain within-study variation and effect size-specific differences among studies. Indeed, the Q test is known to have low power (Sutton and Higgins 2008). A more reliable estimate is the I^2 statistic, which describes the proportion of effect size variation that is due to heterogeneity (Higgins et al. 2003). The I^2 accounts for 81.6% and thus heterogeneity can be classified as "considerable" (Higgins and Green 2008). This result reinforces the findings of the Q test.[15]

For a first impression of the overall relation between CEP and CFP, we calculate the random effects mean effect size without controlling for the specific sources of heterogeneity. In this simple univariate calculation, the identified heterogeneity is integrated by taking into account the between-study variance as part of the weights. This reveals a positive effect size of 0.073[16], which is statistically significant at any conventional level ($p < 0.001$) and close to the finding by Endrikat et al. (2014), who report a statistically significant overall mean effect of 0.083 in the analysis of partial correlations.

In the next step, we check the presence of publication selection bias which could cause distorted mean effects (Stanley and Doucouliagos 2010). As theory provides reasons for a negative, neutral, and positive relation, we have no initial suspicion that the CEP-CFP literature suffers from publication selection. For a graphical test we check the funnel plot[17] and for a statistical test we conduct Egger's regression test[18]. From the simple graphical analysis of the funnel plot, which shows the inverse standard errors against the partial correlations, no asymmetry is visible in the reported effect sizes. The Egger test (Equation (1) without additional Z variables) proves this assumption as it indicates no remarkable dependency of the effects sizes with their standard errors ($t = 0.65$; $p = 0.1612$).

Multiple meta-regression results

The main results are reported in Table 4. All models are based on a WLS approach, as the Breusch-Pagan test (Breusch and Pagan 1979) confirms heteroscedasticity with a test statistic of 210.59, which is statistically significant at any conventional level. The dependent variable in each baseline model is the partial correlation coefficient measuring the intensity of the CEP-CFP relation. Besides the baseline models (1) to (3), we conduct further robustness tests. First, we solely analyze studies considering endogeneity (model 4) and studies published in refereed journals (model 5). Second, we re-run the analysis on semi-partial correlations (model 6) as well as on the same categorical variable as used in Horváthová (2010) (model 7). Furthermore, we compute several alternative MRA model specifications with differing weighting schemes (Appendix L) and the results for Fisher's z-transformed values as effect size to accommodate potential mathematical downsides of the partial correlation coefficients (Appendix M). Finally, the MRA is performed with alternative estimation methods of the between-study variance τ^2, excluding outliers, and using study averages to handle within-study dependencies (Appendix N). In summary, all robustness tests confirm the general findings of the baseline models reported in Table 4.

Referring to Table 4, the absence of publication bias identified by the Egger test can be confirmed again by the insignificant MRA coefficient for the standard error of the effect size SE(ES). As indicated by model (5), this finding even holds when we exclude grey literature. The INTERCEPT value of 0.273 ($t = 2.69$; $p = 0.008$) in column (1), which reveals a positive overall effect of CEP on CFP, must be interpreted with respect to the explanatory variables (this implies that all dummy variables are set to 0 and average values are set for the continuous variables). Thus, there is not one single overall effect. A detailed interpretation of the INTERCEPT value for different combinations of the explanatory variables is subject of the discussion in the subsequent section.

Table 4: Results of the meta-regression analysis

Dep. variable			ES r										ES r_{SP}		ES categorical	
			(1) Random effects cluster robust (Baseline)		(2) Random effects cluster robust (Baseline version 2)		(3) Random effects cluster robust (Baseline version 3)		(4) Random effects cluster robust (Endogeneity considered)		(5) Random effects cluster robust (Only published studies)		(6) Random effects cluster robust		(7) Ordered probit model	
Hyp.	Explanatory variable	*Hyp. sign*	β	SE_{β}	β	SE_{β}	β	SE_{β}	β	SE_{β}	β	SE_{β}	β	SE_{β}	β	SE_{β}
	INTERCEPT	?	0.273	0.101***	0.816	0.184***	0.648	0.135***	0.027	0.089	0.275	0.103***	0.360	0.039***	-	-
	SE(ES)	0	-0.366	0.296	-0.559	0.328*	-0.554	0.327*	-0.330	0.527	-0.404	0.309	-1.738	1.642	-	-
(*H1*)	DVLP	+	0.274	0.083***	-	-	-	-	0.096	0.095	0.278	0.084***	0.046	0.025*	0.073	0.733
	G8	+	-0.027	0.027	-	-	-	-	-0.072	0.054	-0.020	0.029	0.003	0.017	-0.378	0.232
	BRICS	-	-0.218	0.062***	-	-	-	-	-0.304	0.118**	-0.217	0.057***	-0.063	0.029**	-1.334	0.337***
	ASIA	-	-0.157	0.059***	-	-	-	-	-	-	-0.163	0.065**	-	-	0.394	0.718
	EU	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	AFRICA	-	0.109	0.080	-	-	-	-	-	-	0.119	0.076	0.106	0.023***	-0.152	0.505
	US	+	0.086	0.027***	-	-	-	-	0.098	0.039**	0.084	0.029***	0.034	0.015**	0.324	0.219
	GDP	-	-	-	-0.037	0.016**	-	-	-	-	-	-	-	-	-	-
	GDP^2	-	-	-	-	-	-0.002	0.001**	-	-	-	-	-	-	-	-
	GDPGROW	-	-	-	-0.711	0.243***	-0.728	0.245***	-	-	-	-	-	-	-	-
	COMMON	+	-	-	-0.061	0.057	-0.057	0.057	-	-	-	-	-	-	-	-
	CIVIL	-	-	-	-0.148	0.058**	-0.146	0.058**	-	-	-	-	-	-	-	-
(*H2*)	SMALL	-	0.022	0.024	0.061	0.030**	0.062	0.030**	0.057	0.034*	0.011	0.026	0.004	0.015	0.600	0.190***
	MANUF	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SERV	-	-0.069	0.024***	-0.051	0.024**	-0.051	0.024**	-0.112	0.036***	-0.060	0.025**	-0.024	0.017	0.032	0.193
	DIRT	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
(*H3*)	PROCESS-BASED	?	-0.042	0.018**	-0.059	0.017***	-0.059	0.017***	-0.006	0.031	-0.050	0.019***	-0.028	0.013**	-0.099	0.136
	MARKET-BASED	+	0.035	0.019*	0.050	0.020**	0.050	0.020**	0.022	0.026	0.028	0.019	0.011	0.012	0.286	0.163*
	PROACT	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	REACT	-	-0.125	0.034***	-0.116	0.028***	-0.116	0.028***	-0.068	0.044	-0.126	0.034***	-0.057	0.019***	-0.651	0.291**
	CAUSAL	?	0.016	0.013	0.017	0.015	0.017	0.015	0.008	0.016	0.016	0.014	0.014	0.011	0.162	0.071**
(*H4*)	MILLENIUM	?	-0.298	0.070***	-0.324	0.071***	-0.318	0.071***	-0.101	0.079	-0.271	0.073***	-0.373	0.038***	-0.460	0.434
	DATERANGE	?	0.003	0.003	0.005	0.003*	0.005	0.003*	0.004	0.006	0.004	0.003	-0.004	0.002**	0.062	0.019***
	NOFIRMS	?	-0.000	0.000**	-0.000	0.000**	-0.000	0.000**	-0.000	0.000	-0.000	0.000*	-0.000	0.000	-0.000	0.000
	1980s	?	-0.240	0.066***	-0.239	0.068***	-0.236	0.068***	-	-	-0.250	0.069***	-0.300	0.035***	-0.257	0.379
	1990s	?	-0.172	0.062***	-0.190	0.064***	-0.185	0.065***	0.031	0.049	-0.192	0.064***	-0.268	0.030***	-0.270	0.324
	2000s	?	0.111	0.035***	0.142	0.041***	0.143	0.041***	0.094	0.067	0.070	0.039*	0.116	0.021***	0.367	0.289
	MLTNAT	?	0.125	0.023***	-0.015	0.049	-0.015	0.049	0.012	0.070	0.130	0.026***	0.065	0.016***	0.164	0.247
	NOCOUNT	?	-0.010	0.002***	-0.009	0.002***	-0.009	0.002***	-0.006	0.004	-0.010	0.002***	-0.006	0.001***	-0.052	0.016***

MLTSECT	?	0.025	0.022	0.014	0.027	0.013	0.027	0.071	0.036*	0.029	0.023	-0.014	0.016	-0.144	0.190
NOSECT	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NOAUTHORS	?	0.040	0.009***	0.031	0.009***	0.031	0.009***	0.048	0.012***	0.041	0.010***	0.015	0.005***	0.183	0.060***
CITE	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SMPL	?	0.030	0.027	0.028	0.027	0.027	0.027	0.052	0.038	0.035	0.032	-0.000	0.026	0.517	0.211**
HTRG	?	-0.017	0.031	-0.020	0.032	-0.020	0.032	-0.053	0.044	-0.034	0.035	0.002	0.025	-0.630	0.233***
ENDO	?	-0.024	0.020	-0.010	0.020	-0.009	0.020	-	-	-0.016	0.023	-0.016	0.016	-0.057	0.158
RD	?	0.024	0.022	0.018	0.023	0.018	0.023	0.013	0.037	0.003	0.031	0.019	0.021	-0.055	0.186
AD	?	0.015	0.026	0.023	0.026	0.023	0.026	0.038	0.054	0.018	0.030	-0.013	0.022	0.196	0.245
CAP	?	-0.085	0.020***	-0.082	0.020***	-0.081	0.020***	-0.033	0.029	-0.067	0.027**	-0.056	0.017***	-0.286	0.180
RISK	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GROW	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SIZE	?	-0.050	0.022**	-0.045	0.024*	-0.044	0.024*	-0.045	0.033	-0.040	0.023*	-0.047	0.013***	0.032	0.158
IND	?	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AIC		2,512.99		2,479.5		2,479.42		1,077.38		2,250.09		1,961.77		-	
BIC		2,660.56		2,617.30		2,617.17		1,182.68		2,394.25		2,097.35		-	
τ^2		0.015		0.016		0.016		0.018		0.016		0.011		-	
Adj. R^2		0.269		0.252		0.252		0.212		0.250		0.179		0.115	
Observations		863		854		854		365		773		678		863	
No. of studies		133		131		131		66		124		108		133	

This table presents the final results of the meta-regression model stated in Equation (1). The dependent variable in model (1) through model (5) is the intensity of the CEP-CFP relation measured by the partial correlation coefficient *r* as effect size. As a robustness test we use the semi-partial correlation as proposed by Aloe and Becker (2012) in model (6) in order to correct for the application of multiple regression models in primary studies. Further, the vote counting variable in model (7) as applied by Horváthová (2010) indicating a significant positive (=1), insignificant (=0) or significant negative (=-1) result based on the 5% level. The explanatory variables are grouped to test the proposed hypotheses *H1* through *H4* and measure different study characteristics. A detailed description of the variables can be found in Section 4 of the paper. Model (1) through model (6) further use weighted least squares to accommodate heteroscedasticity, standard errors are adjusted for study level data dependencies by using cluster robust errors. Moreover, the random effects are estimated by the restricted maximum-likelihood estimator (REML), τ^2 is the corresponding estimate for the between-study variance. All models report parsimonious models based on general-to-specific modeling. Therein, after each re-estimation the most insignificant variable is dropped from the model until the remaining test statistics are consistently larger than 1. Model (2) and (3) report the results for the alternative country-specific variables. Model (4) is calculated on the sub-sample of studies that correct their results for potential endogeneity between CEP and CFP. Model (5) is solely based on effect sizes from published studies.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Across all tested explanatory variables, the results show that several factors significantly influence the reported effect between CEP and CFP. Moreover, the heterogeneity arises not only from differences in the study design, but especially from real factors such as differences among regions, industries, as well as from differences in the measurement of CEP and CFP.

Regional differences

Concerning the CEP-CFP estimates for differences in economic development and geographical regions, four out of the seven explanatory factors are statistically significant at the 1% level (DVLP, BRICS, ASIA, US). With a positive and significant regression coefficient of 0.274 for DVLP, the CEP-CFP relation is (on average) amplified in developing countries. This finding might be a first indicator that the CEP-CFP link depends on the level of economic development. Ferron et al. (2012) suggest that enhanced efficiency by saving materials and preventing waste highly affects the cost structure in these countries. Arafat et al. (2012), who state that enhanced CEP affects learning in developing countries, confirm this result. Further, Agan et al. (2014) propose that the higher CEP response to CFP in developing countries may also be explained by environmental supplier development. Hence, international firms might enhance their CEP by investing in suppliers located in developing countries. With a highly significant coefficient of -0.218, the BRICS variable implies that reported estimates for the CEP-CFP relation are diminished in BRICS countries. Dinda (2004) states that in high growth countries, economic success is preferred against environmental responsibility. The ASIA variable confirms this result by a highly significant coefficient with a value of -0.157. Furthermore, studies examining US firms seem to yield systematically different results than non-US studies, which is in line with Dixon-Fowler et al. (2013) as well as Horváthová (2010), but contradicting to Albertini (2013). Thus, better CEP coincides more often with better CFP for US firms than for the rest of the world. Albertini (2013) as well as Jaffe and Palmer (1997) notice that this might be explained by the combination of flexible environmental policies and high profitability. Finally, looking at the G8 countries and the EU, we find no systematic differences. To provide more insights into the impact of regional differences on the reported estimates for the CEP-CFP nexus, we add further country-specific variables. These refer to an alternative measurement of the economic development and regulatory environment. The results are reported in models (2) and

(3) of Table 4. The variables GDP, GDPGROW, and CIVIL[19] are all statistically significant, at least at the 5% level. Thus, we can conclude that macro-economic, legal, and governance factors seem to systematically affect the estimated relation between CEP and CFP. In more detail, the finding for GDP with a MRA coefficient of -0.037 suggests that CEP-CFP effect weakens for economically more developed countries. In addition, the phase of the economic cycle measured by GDPGROW explains heterogeneity in reported outcomes. Our results for GDP and GDPGROW are accompanied by the finding of Di Vita (2009), who reveals a positive relation between pollution level and per capita income as well as economic growth rate. We additionally inspect the results for a model specification with a squared GDP[20] variable. The resulting coefficient with a value of -0.002 is statistically significant at the 5% level. Moreover, the results indicate that the reported partial correlations are, on average, diminished in civil law countries. This result is line with Horváthová (2010) and Di Vita (2009), who propose that the positive link between CEP and CFP more often occurs in common law countries that offer greater protection to shareholders and creditors.

Industrial differences

Among industrial differences, the SERV variable with a coefficient of -0.069 appears to be systematically important for explaining differences in the reported CEP-CFP effect sizes. As Inoue and Lee (2011) as well as Villiers et al. (2011) propose that the negative coefficient suggests that service sector firms less effectively deal with environmental performance compared to dirtier industries. Kassinis and Soteriou (2005) argue that financial returns, especially from enhancement of customer satisfaction and loyalty through better CEP, are smaller for firms from the service industry compared to other industries, which is due to weaker differentiation possibilities through "green" behavior. For the SMALL variable, only the results with the alternative country-specific variables indicate explanatory power for this factor. This speaks for a larger effect size for samples that are restricted to small enterprises. The variables for manufacturing firms and dirty industries do not seem to yield systematically differences in the reported results. This is in line with the findings by Albertini (2013).

Differences in variable operationalization

Among the factors capturing differences in variable operationalization, two variables condition the relation between CEP and CFP. First, PROCESS-BASED has a significant regression coefficient of -0.042, which shows that the effect between CEP and CFP is amplified for outcome-based compared to process-based CEP measures. This contradicts the finding by Horváthová (2010). Busch and Hoffmann (2011) point out that stakeholders favor valuing CEP based on outcome-based measures compared to process-based measures. Concerning this circumstance, Xie and Hayase (2007) explain that the environmental process improvements as captured by process-based measures do not necessarily end up in better CEP. Furthermore, Delmas et al. (2013) and Wagner (2008) evaluate process-based measures as more subjective and less reliable for stakeholders. Second, the REACT variable shows a negative sign (-0.125) and is highly significant at the 1% level. Following, for example, primary studies by King and Lenox (2002) as well as Klassen and Whybark (1999) and meta-analyses by Endrikat et al. (2014), Guenther et al. (2012), as well as Murphy (2002), reactive investments are confirmed to coincide less often with enhanced CFP compared to proactive investments. Proactive activities as environmental management strategy induce new technologies and capabilities of employees that lead to strategic and therefore financial advantages for the firm (Hart 1995; Russo and Fouts 1997). Finally, compared to the previous work by Albertini (2013), the study only finds a contradicting and weakly significant positive effect for the MARKET-BASED variable. According to Dixon-Fowler et al. (2013), this result speaks for a larger effect size reported in studies using market-based CFP measures. Thus, future-oriented market-based measures indeed reveal the positive expectations from shareholders and long-run benefits from better environmental performance.

Further moderators

Several aspects of data characteristics seem to be highly relevant for the explanation of varying effect sizes. The results for MILLENIUM, 1980s, 1990s, and 2000s demonstrate temporal deviations of the relation between CEP and CFP. Furthermore, the negative sign of the NOFIRMS variable suggests that larger sample studies report, on average, smaller effects. In addition, multinational samples yield higher effect sizes than single country examinations as confirmed by the positive and highly significant result for MLTNAT. Moreover, the assumption that the link between CEP and CFP blurs with an increasing number of countries (Berman et al. 1999) is verified by the NOCOUNT variable, which has a statistically significant value of -0.010. Additionally, among the publication characteristics the number of authors partly explains the heterogeneity with coefficient of 0.040 and a corresponding *p*-value below 0.001. Accordingly, this result indicates that larger research teams report, on average, amplify the CEP-CFP relation. However, one should be cautious with the interpretation of this variable due to the weak theoretical foundation.

We find no evidence that estimation differences are responsible for deviations in the reported effects for the relationship between CEP and CFP. Concerning the crucial issue of endogeneity, our results of model (4) reveal that regional and industrial effects remain even while accounting for endogeneity between CEP and CFP. However, measurement differences and temporal effects seem to be diminished. The latter might be caused by the fact that especially recent studies consider endogeneity, which biases the estimation of time effects.

As a special capability of MRA, we account for potential misspecification biases in reported effect estimates caused by omitted variables in existing primary research (Stanley 2001). We find capital intensity (CAP) as well as firm size (SIZE) to be significant (at least at the 1% level and 5% level respectively) with a regression coefficient of -0.085 and -0.050 respectively. This confirms the relevance of these control variables for the investigation of the CEP-CFP relation. The negative sign for CAP indicates that the omission of capital intensity as control variable is related with a larger reported effect size. Further, the significantly negative sign of SIZE in model (1) confirms that models which control for firm size report partial correlations that are, on average, -0.05 lower compared to models omitting this variable.

Although the majority of robustness tests lead to the same findings, several deviations are remarkable. For model (4), the disappeared time effects (*H4a*) might be driven by the fact that especially recent studies consider endogeneity and older studies are ignored in this sub-sample. Furthermore, our results let assume that deviations in the results from measurement differences

(*H3*) might be diminished by accounting for endogeneity in the model estimation. As shown in model (6), a vote counting approach leads to different results for estimation differences (*H4c*) and model specifications (*H4d*). As the coefficients for the variables testing the two hypotheses have small values, these marginal coefficients are not identified using the more generic vote counting approach.

Discussion

For practical implications of the results from the multiple MRA, we calculate mean effects for CEP-CFP by substituting plausible values for the explanatory variables (Linde Leonard et al. 2014). As an advantage over univariate analyses, this approach allows us to investigate results for the CEP-CFP relationship in a multidimensional manner from "best practice research" meaning an advanced and recent study. Our findings show that these results are robust against several methodological changes in the "best practices", which is also driven by the small values and thus tiny impact of many variables from *H2, H3,* and *H4*.

Looking at our baseline model (1) from Table 4, we plug in the value 0 for SE. This means that publication bias is eliminated and that we have very precise estimates from a large sample ($n \to \infty$; $SE \to 0$). Continuing with the Z vector in the same manner, we include the significant control variables CAP and SIZE. Further, we assume a sophisticated statistical approach, considering unobserved heterogeneity as well as endogeneity for the estimation characteristics. For the publication and data characteristics we plug in the mean values, except for the time variables. The latter are set to represent a recent study using data from the 2000s. For the measurement differences, we assume a concurrent measurement of CEP and CFP, a proactive investment, an accounting-based CFP measure, and an outcome-based CEP measure. For the industry, regional differences, as well as firm size, we insert several alternative variable combinations as shown in Table 5. For an exemplary calculation of the mean effect reported in the first row of Table 5, please refer to the corresponding endnote[21].

Following the guidelines for the interpretation of partial correlation coefficients by Doucouliagos (2011)[22], the value for Africa, which is 0.427, speaks for a large effect in developing countries. This effect is nearly unaffected by the other integrated variables. The large value remains even when changing the variables for the industrial and measurement characteristics. For the large positive effect in developing countries we especially suspect enhanced efficiency by saving materials and preventing waste (Ferron et al. 2012), learning effects (Arafat et al. 2012), and environmental supplier development (Agan et al. 2014) as the most important factors in developing countries. This also provides conditional evidence for the NRBV. In contrast to developing countries, the effect in emerging or transition countries, which is represented by the BRICS variable, is clearly negative with a value of -0.174. This value can be interpreted as a moderate effect according to the guidelines by Doucouliagos (2011) and is an indication towards the managerial opportunism hypothesis in high-growth countries where CFP is the major priority. When the level of development further increases, the CEP-CFP relation becomes again positive, which is shown by the mean effect of 0.103 for US firms. Thus, in summary, we observe a convex relationship between the economic development of a country and the CEP-CFP relationship. This means that the appreciation of environmental responsibility coincides with a certain level of prosperity (Dinda 2004). Remembering the transition of the US (The Washington Post 2014) from a manufacturing economy to a service economy, this positive effect is diminished to only 0.012 and thus almost zero. Finally, in this context it should be noted that complying with regulations by reactive investments consequently produces a negative sign for the CEP-CFP relation. With a value of -0.125, this might have a practical impact, which makes proactive strategies to the superior choice. This might also provide an incentive for policy makers to create political conditions where “it pays to be green”, which would be in line with the Porter hypothesis.

Table 5: Estimates for best practice

Country	Industry	Firm size	Estimate for CEP-CFP relation
Africa	Manufacturing sector	Small	0.427
BRICS	Manufacturing sector	Small	-0.174
US	Manufacturing sector	Small	0.103
US	Manufacturing sector	Large	0.081
US	Service sector	Large	0.012

This table presents the estimated effect sizes from best practice estimates for several variable combinations. The country, industry, and firm size variable is varied. The estimated effect sizes are presented in the last column.

Summary of results and implications

Compared to previous univariate findings, our MRA results provide new insights into the multidimensional determinants of the CEP-CFP relation. A summary of the overall findings for the main MRA variables can be found in Table 6.

Table 6: Summary of key results

Variable	Sign.	Finding	Inference drawn from
Robust findings			
Regional differences (*H1*)			
DVLP	Strong	For developing countries the reported effects between CEP and CFP are, on average, higher.	Model 1
BRICS	Strong	The reported effects between CEP and CFP are, on average, smaller for BRICS countries.	Model 1
ASIA	Strong	For Asian countries the reported effects between CEP and CFP are, on average, smaller.	Model 1
US	Strong	Reported effects between CEP and CFP are, on average, larger for the US.	Model 1
GDP	Strong	The reported effects between CEP and CFP are, on average, smaller for countries with a large GDP per capita, which also holds for the squared factor.	Model 2 and 3
CIVIL	Strong	For civil law countries, on average, reported effects between CEP and CFP are smaller.	Model 2 and 3
Differences in variable operationalization (*H3*)			
PROCESS-BASED	Weak	Outcome-based CEP measures, on average, lead to higher reported effects between CEP and CFP than process-based measures.	Model 1
REACT	Strong	On average, reactive environmental activities lead to lower reported effects.	Model 1
Other findings			
Industrial differences (*H2*)			
SERV	Strong	For the service sector, the reported effects between CEP and CFP is, on average, smaller.	Model 1
Differences in variable operationalization (*H3*)			
MARKET-BASED	Weak	Market-based CFP measures lead, on average, to higher reported effects between CEP and CFP than accounting-based measures.	Model 1

This table sums up the results from random effects weighted least squares MRA. The "robust findings" are based on their consistent significance in the reported results in Table 4. The "other findings" are drawn solely from the preferred models (1) through (3). Beside the respective explanatory variable, the significance of its effect on the CEP-CFP relation as well as a short interpreting statement are given. The last column shows where the finding is located in the results table. Strong significance refers to the 99% level and weak significance the 95% level.

There a several main conclusions to be drawn from the MRA approach:

(1) We examine so far rarely investigated factors and find that especially differences in a country's economic development and legal system matter for the explanation of the heterogeneity in the empirical research on the CEP-CFP relation.

(2) There is only weak support that industrial characteristics are responsible for the heterogeneous results.

(3) For the measurement of the variables, we confirm previous findings which show that the choice of process-based measures, reactive environmental activities, and market-based performance figures have an impact on the reported outcomes.

(4) Regarding the further examined variables, our results reveal that the examined time period, the sample size, as well as the inclusion of control variables for capital intensity and firm size are responsible for differences in the primary studies.

These results are also relevant for international management behavior and for policy. For the group of corporate decision makers, it is important to consider the geographical region when deciding about environmental investments. Moreover, they should avoid reactive environmental activities. Policy makers should care about the systematic differences in the CEP-CFP relationship across differing economic developments and legal systems. Finally, these implications are interesting for environmental actions on the supranational level, as for example the European Union or the United Nations Climate Change Conferences.

6. Conclusion

By extending previous reviews and meta-analyses (Albertini 2013; Dixon-Fowler et al. 2013; Endrikat et al. 2014; Guenther et al. 2012; Horváthová 2010; Murphy 2002), this paper aims at explaining the heterogeneity of existing empirical results on the CEP-CFP relation. We use a large set of 42 different explanatory variables to measure the impact of regional and industrial differences as well as several other dimensions of research design. Thereby, we directly build on the theoretical framework of moderating and mediating factors developed by Guenther and Hoppe (2014), who ask to utilize the existing empirical findings and focus on the empirical identification of the conditions that foster the prospects of success for corporate environmental activities. We address the multivariate dependence-structure among the explanatory variables by simultaneously testing the relevant explanatory variables in a multiple MRA.

In summary, we find that deviations from the overall positive relation between CEP and CFP are especially driven by the geographic location and economic performance of a country, differences among industry sectors, the type of the environmental activity and the measurement of the financial performance. In detail, the results suggest a convex impact of a country's economic development on the magnitude of the CEP-CFP effect, i.e. the effect is positive in developing countries, disappears in emerging countries, and is again positive in highly developed countries.

This supports the view that the CEP-CFP relation behaves like the environmental Kuznets curve. Moreover, the results reveal that the overall positive relation strengthens for market-based CFP measures and diminishes for countries with civil law systems, firms from the service sector, reactive environmental activities, and process-based CEP measures. Furthermore, several aspects of the examined data sample and the inclusion of relevant control variables explain heterogeneity in previous research results. These findings should encourage decision-makers to invest in proactive environmental activities, all the while considering the location and industrial conditions of the firm.

Nevertheless, our study is limited by several issues and raises questions for future research. First, several potentially moderating factors presented by Guenther and Hoppe (2014) are not considered in this article, as firm-specific data are necessary which are indeed not available on a meta-level and thus should be tested in further research on the level of primary studies. The potentially omitted aspects encompass the stakeholder type (for example, Moore 2001), rivals' cost (for example, McWilliams et al. 2002), firm risk (for example, Sharfman and Fernando 2008), and quality of management (for example, Corbett et al. 2005). Second, as we feel this is an interesting area, one might further analyze cultural factors, measures of globalization, as well as additional industry-level characteristics to obtain a more in-depth understanding of the regional and industrial conditions. Especially, the theoretical background behind the observed effect between the economic development of a country and the CEP-CFP relation is still open. Third, we suggest applying a meta-analytic structural equation model (Cheung and Chan 2005) for a deeper analysis of the mediating factors of the CEP-CFP relation (Guenther and Hoppe 2014). Finally, our study proposes that endogeneity is a fundamental issue in empirical research on the CEP-CFP relation by revealing an impact on several explanatory factors. Thus, we encourage applying appropriate estimation techniques to receive meaningful results (Bascle 2008; Hamilton and Nickerson 2003). In summary, the endogeneity between CEP and CFP is still pending and the assessment of the appropriateness of these endogeneity treatments should be subject of further research. Until then, one should take care of the interpretation of causal mechanisms underlying the CEP-CFP nexus.

Supplementary material

All appendices connected to this paper are available in the corresponding e-component. This supporting information includes a reference list for the meta-regression analysis, operationalization of the utilized explanatory variables, the funnel plot of the CEP-CFP relation, and ten additional tables. The tables show reporting guidelines for meta-regression research, an overview of the literature search process in the electronic databases, a detailed overview of the study sample, publication outlet of papers included in the meta-analysis, temporal distribution of primary studies, formulas for the calculation of effect sizes, univariate meta-analytical results, results of alternative MRA models, results of *z*-transformed effect sizes, and further robustness tests.

References[23]

(The citations marked with an asterisk * are solely cited in the e-component.)

Abdullah, A., H. Doucouliagos, and E. Manning. 2015. Does Education Reduce Income Inequality? A Meta-Regression Analysis. *Journal of Economic Surveys*, 29(2): 301–316.

Agan, Y., C. Kuzey, M. F. Acar, and A. Açıkgoez. 2014. The relationships between corporate social responsibility, environmental supplier development, and firm performance. *Journal of Cleaner Production*, 51: 23–33.

Albertini, E. 2013. Does environmental management improve financial performance? A meta-analytical review. *Organization & Environment*, 26(4): 431–457.

Allouche, J., and P. Laroche. 2005. A meta-analytical investigation of the relationship between corporate social and financial performance. *Revue de gestion des ressources humaines*, 57: 18–41.

*Alkhafaji, A. F. 1989. *A stakeholder approach to corporate governance: Managing in a dynamic environment.* Westport: Greenwood Press.

Ankamah-Yeboah, I. and K. Rehdanz. 2014. *Explaining the variation in the value of building energy efficiency certificates: A quantitative meta-analysis.* Working paper, Kiel Institute for the World Economy (IfW), Kiel, Germany.

Arafat, M. Y., A. Warokka, and S. R. Dewi. 2012. Does environmental performance really matter? A lesson from the debate of environmental disclosure and firm performance. *Journal of Organizational Management Studies*, 2012: 15 pages.

Aragon-Correa, J. A., A. Marcus, and N. Hurtado-Torres. 2015. The Natural Environmental Strategies of International Firms: Controversies and New Evidence on Performance and Disclosure. *The Academy of Management Perspectives*, 30(1): 24–39.

Barnett, M. L. 2007. Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *The Academy of Management Review*, 32(3): 794–816.

Barney, J. 1991. Firm resources and sustained competitive advantage. *Journal of Management*, 17(1): 99–120.

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

Bausch, A., and F. Pils. 2009. Product diversification strategy and financial performance: meta-analytic evidence on causality and construct multidimensionality. *Review of Managerial Science*, 3(3): 157–190.

Begg, C. B., and J. A. Berlin. 1988. Publication bias: a problem in interpreting medical data. Journal of the Royal Statistical Society. *Series A,* 151: 419–463.

Bender, R., C. Bunce, M. Clarke, S. Gates, S. Lange, N. L. Pace, and K. Thorlund. 2008. Attention should be given to multiplicity issues in systematic reviews. *Journal of Clinical Epidemiology*, 61(9): 857–865.

Berman, S. L., A. C. Wicks, S. Kotha, and T. M. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *The Academy of Management Journal*, 42(5): 488–506.

Bernauer, T. and T. Boehmelt. 2013. Are Economically "Kinder, Gentler Societies" also Greener? *Environmental Science & Technology*, 47: 11993–12001.

Berrone, P. and L. R. Gomez-Mejia. 2009. Environmental performance and executive compensation: An integrated agency-institutional perspective. *The Academy of Management Journal*, 52(1): 103–126.

Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009. *Introduction to Meta-Analysis.* Chichester: John Wiley and Sons.

Bourgeois, L. J. 1981. On the measurement of organizational slack. *The Academy of Management Review*, 6(1): 29–39.

Bragdon Jr, J. H. and J. A. Marlin. 1972. Is Pollution Profitable. *Risk Management*, 19: 9–18.

*Branco, M. C. and L. L. Rodrigues. 2006. Corporate social responsibility and resource-based perspectives. *Journal of Business Ethics*, 69(2): 111–132.

Breusch, T. S. and A. R. Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5): 1287–1294.

Brouwers, R., F. Schoubben, C. van Hulle, and S. van Uytbergen. 2014. The link between corporate environmental performance and corporate value: A literature review. *Review of Business and Economic Literature*, 58(4): 343–374.

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

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

Card, D. and A. B. Krueger. 1995. Time-series minimum-wage studies: a meta-analysis. *The American Economic Review*, 85(2): 238–243.

Carney, M., E. R. Gedajlovic, P. P. Heugens, M. van Essen, and J. H. van Oosterhout. 2011. Business group affiliation, performance, context, and strategy: A meta-analysis. *Academy of Management Journal*, 54(3): 437–460.

Chaikumbung, M., C. Doucouliagos, and H. Scarborough. 2016. The economic value of wetlands in developing countries: A meta-regression analysis. *Ecological Economics*, 124: 164–174.

Cheung, M. W.-L. and W. Chan. 2005. Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, 10(1): 40–64.

Cochran, W. G. 1954. The combination of estimates from different experiments. *Biometrics*, 10(1): 101–129.

Cohen, M. A., S. A. Fenn, and S. Konar. 1997. *Environmental and financial performance: Are they related?* Working paper, Vanderbilt University, Nashville, USA.

Corbett, C. J., M. J. Montes-Sancho, and D. A. Kirsch. 2005. The Financial Impact of ISO 9000 Certification in the United States: An Empirical Analysis. *Management Science*, 51(7): 1046–1059.

*Cornell, B. and A. C. Shapiro. 1987. Corporate stakeholders and corporate finance. *Financial Management*, 16: 5–14.

*Damak-Ayadi, S. and Y. Pesqueux. 2005. Stakeholder theory in perspective. *Corporate Governance: The International Journal of Business in Society*, 5(2): 5–21.

Davis, K. 1973. The case for and against business assumption of social responsibilities. *The Academy of Management Journal*, 16(2): 312–322.

Day, R., A. Amati, and B. Neubert. 1997. *The financial impact of environmental events and issues on the forest products industr*y. Office of Policy, Planning and Evaluation, Washington, D.C., USA.

Delmas, M. A. and N. S. Nairn-Birch. 2011. *Is the tail wagging the dog? An empirical analysis of corporate carbon footprints and financial performance.* Working paper, University of California, Los Angeles, USA.

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

Delmas, M., V. H. Hoffmann, and M. Kuss. 2011. Under the tip of the iceberg: Absorptive capacity, environmental strategy, and competitive advantage. *Business & Society*, 50(1): 116–154.

Demena, B. A. and P. A. van Bergeijk. 2016. A meta-analysis of FDI and productivity spillovers in developing countries. *Journal of Economic Surveys.* In press.

Dinda, S. 2004. Environmental Kuznets curve hypothesis: a survey. *Ecological Economics*, 49(4): 431–455.

Di Vita, G. 2009. Legal families and environmental protection: Is there a causal relationship? *Journal of Policy Modeling*, 31(5): 694–707.

Dixon-Fowler, H. R., D. J. Slater, J. L. Johnson, A. E. Ellstrand, and A. M. Romi. 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.

Doucouliagos, H. 2011. *How Large is Large? Preliminary and relative guidelines for interpreting partial correlations in economics*. Working Paper, Deakin University, Melbourne, Australia.

Doucouliagos, H. and P. Laroche. 2009. Unions and Profits: A Meta-Regression Analysis. *Industrial Relations: A Journal of Economy and Society*, *48*(1): 146–184.

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

Doucouliagos, H. and T. D. Stanley. 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2): 406–428.

Doucouliagos, C. and T. D. Stanley. 2013. Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, 27(2): 316–339.

Doucouliagos, H. and M. A. Ulubasoglu. 2008. Democracy and economic growth: A meta-analysis. *American Journal of Political Science*, 52(1): 61–83.

Egger, M., G. D. Smith, M. Schneider, and C. Minder. 1997. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109): 629–634.

Elsayed, K. and D. Paton. 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., E. M. Guenther, and H. Hoppe. 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.

Esty, D. C. and M. E. Porter. 1998. Industrial ecology and competitiveness. *Journal of Industrial Ecology*, 2(1): 35–43.

Ferron, R. T., B. Funchal, V. Nossa, and A. J. C. Teixeira. 2012. Is ISO 14001 certification effective? An experimental analysis of firm profitability. *BAR-Brazilian Administration Review*, 9: 78–94.

García-Quevedo, J. 2004. Do public subsidies complement business R&D? A meta-analysis of the econometric evidence. *Kyklos*, 57(1): 87–102.

Ginsberg, A. 1988. Measuring and modelling changes in strategy: Theoretical foundations and empirical directions. *Strategic Management Journal*, 9(6): 559–575.

Goerg, H. and E. Strobl. 2001. Multinational companies and productivity spillovers: a meta-analysis. *Economic Journal*, 111(475): 723–739.

Greene, W. H. 2011. *Econometric analysis*. Seventh edition. London: Prentice Hall.

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

Gurdgiev, C., B. M. Lucey, and D. Reilly. 2010. *Real Estate and the Stock Market: A Meta-Regression Analysis*. Working paper, Trinity College Dublin, Dublin, Ireland.

Guenther, E. M. and H. Hoppe. 2014. Merging Limited Perspectives. *Journal of Industrial Ecology*, 18(5): 689–707.

Guenther, E. M., H. Hoppe, and J. Endrikat. 2012. Corporate financial performance and corporate environmental performance: A perfect match? *Zeitschrift für Umweltpolitik und Umweltrecht*, 34: 279–296.

Hamilton, B. H. and J. A. Nickerson. 2003. Correcting for endogeneity in strategic management research. *Strategic organization*, 1(1): 51–78.

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

Havránek, T. 2010. Rose effect and the euro: is the magic gone? *Review of World Economics*, 146(2): 241–261.

Havránek, T., R. Horvath, Z. Irsova, and M. Rusnak. 2015. Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, 96(1): 100–118.

Havránek, T. and Z. Irsova. 2011. Estimating vertical spillovers from FDI: Why results vary and what the true effect is. *Journal of International Economics*, 85(2): 234–244.

Hedges, L. V. 1992. Modeling publication selection effects in meta-analysis. *Statistical Science*, 7: 246–255.

Hedges, L. V. and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. London: Academic Press.

Hedges, L. V., E. Tipton, and M. C. Johnson. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, 1(1): 39–65.

Henriques, I. and P. Sadorsky. 1999. The relationship between environmental commitment and managerial perceptions of stakeholder importance. *The Academy of Management Journal*, 42(1): 87–99.

Heras-Saizarbitoria, I., J. F. Molina-Azorín, and G. P. M. Dick. 2011. ISO 14001 certification and financial performance: selection-effect versus treatment-effect. *Journal of Cleaner Production*, 19(1): 1–12.

Higgins, J. P. and S. Green. 2008. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell.

Higgins, J. P. and S. G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11): 1539–1558.

Higgins, J. P., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414): 557–560.

Horváthová, E. 2010. Does environmental performance affect financial performance? A meta-analysis. *Ecological Economics*, 70(1): 52–59.

Hughes, S. B., A. Anderson, and S. Golden. 2001. Corporate environmental disclosures: are they useful in determining environmental performance? *Journal of Accounting and Public Policy*, 20(3): 217–240.

Hunter, J. E. and F. L. Schmidt. 2004. *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage.

Ilinitch, A. Y., N. S. Soderstrom, and T. E. Thomas. 1999. Measuring corporate environmental performance. *Journal of Accounting and Public Policy*, 17(4): 383–408.

Inoue, Y. and S. Lee. 2011. Effects of different dimensions of corporate social responsibility on corporate financial performance in tourism-related industries. *Tourism Management*, 32(4): 790–804.

Iwata, H. and K. Okada. 2011. How does environmental performance affect financial performance? Evidence from Japanese manufacturing firms. *Ecological Economics*, 70(9): 1691–1700.

Jaffe, A. B. and K. Palmer. 1997. Environmental regulation and innovation: A panel data study. *Review of Economics and Statistics*, 79: 610–619.

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

Kassinis, G. I. and A. C. Soteriou. 2003. Greening the service profit chain: The impact of environmental management practices. *Production and Operations Management*, 12(3): 386–403.

Kerret, D. and R. Shvartzvald. 2012. Explaining Differences in the Environmental Performance of Countries: A Comparative Study. *Environmental Science & Technology*, 46: 12329–12336.

King, A. A. and M. J. Lenox. 2001. Does it really pay to be green? An empirical study of firm environmental and financial performance: An empirical study of firm environmental and financial performance. *Journal of Industrial Ecology*, 5(1): 105-116.

King, A. A. and M. J. Lenox. 2002. Exploring the locus of profitable pollution reduction. *Management Science*, 48(2): 289–299.

Klassen, R. D. and D. C. Whybark. 1999. The impact of environmental technologies on manufacturing performance. *The Academy of Management Journal*, 42(6): 599–615.

Kock, C. J., J. Santaló, and L. Diestre. 2012. Corporate governance and the environment: what type of governance creates greener companies? *Journal of Management Studies*, 49(3): 492–514.

Konar, S. and M. A. Cohen. 2001. Does the market value environmental performance? *Review of Economics and Statistics*, 83(2): 281–289.

*Kraaijenbrink, J., J.-C. Spender, and A. J. Groen. 2010. The resource-based view: A review and assessment of its critiques. *Journal of Management*, 36(1): 349–372.

Kraft, K. L. and J. Hage. 1990. Strategy, social responsibility and implementation. *Journal of Business Ethics*, 9(1): 11–19.

Kysucky, V. and L. Norden. (2016). The benefits of relationship lending in a cross-country context: A meta-analysis. *Management Science*, *62*(1): 90–110.

Latané, B. 1981. The psychology of social impact. *American psychologist*, 36(4): 343–358.

Lee, D. and R. Madhavan. 2010. Divestiture and firm performance: A meta-analysis. *Journal of Management*, 36(6): 1345–1371.

Levitt, T. 1958. The dangers of social-responsibility. *Harvard Business Review*, 36(5): 41–50.

Linde Leonard, M., T. D. Stanley, and H. Doucouliagos. 2014. Does the UK Minimum Wage Reduce Employment? A Meta-Regression Analysis. *British Journal of Industrial Relations*, 52(3): 499–520.

Llach, J., J. Perramon, M. del Mar Alonso-Almeida, and L. Bagur-Femenias. 2013. Joint impact of quality and environmental practices on firm performance in small service businesses: an empirical study of restaurants. *Journal of Cleaner Production*, 44: 96–104.

McGrath, D. T. and S. Gilbert-Miller. 1999. Transferring lean and green: An empirical evaluation of the adoption of pollution prevention. *Pollution Prevention Review*, 9(3): 49–66.

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

McWilliams, A. and D. Siegel. 2001. Corporate social responsibility: A theory of the firm perspective. *The Academy of Management Review*, 26(1): 117–127.

McWilliams, A. and D. Siegel, and P. M. Wright. 2006. Corporate social responsibility: Strategic implications. *Journal of Management Studies*, 43(1): 1–18.

McWilliams, A., D. D. van Fleet, and K. D. Cory. 2002. Raising Rivals' Costs Through Political Strategy: An Extension of Resource-based Theory. *Journal of Management Studies*, 39(5): 707–724.

Miller, D. and P. H. Friesen. 1983. Strategy-making and environment: The third link. *Strategic Management Journal*, 4(3): 221–235.

Molina-Azorín, J. F., E. Claver-Cortés, M. D. López-Gamero, and J. J. Tarí. 2009. Green management and financial performance: A literature review. *Management Decision*, 47(7): 1080–1100.

Moore, G. 2001. Corporate social and financial performance: An investigation in the UK supermarket industry. *Journal of Business Ethics*, 34(3-4): 299–315.

Muhammad, N., F. Scrimgeour, K. Reddy, and S. Abidin. 2015. The relationship between environmental performance and financial performance in periods of growth and contraction: evidence from Australian publicly listed companies. *Journal of Cleaner Production*, 102: 324–332.

Murphy, C. J. 2002. *The profitable correlation between environmental and financial performance: A review of the research.* Prepared for Light Green Advisors, New York.

Nelson, J. P. and P. E. Kennedy. 2009. The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental and Resource Economics*, 42(3): 345–377.

Orlitzky, M. 2001. Does firm size confound the relationship between corporate social performance and firm financial performance? *Journal of Business Ethics*, 33(2): 167–180.

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

Patten, D. M. 2002. The relation between environmental performance and environmental disclosure: A research note. *Accounting, Organizations and Society*, 27(8): 763–773.

Porter, M. E. and C. van der Linde. 1995a. Green and competitive: ending the stalemate. *Harvard Business Review*, 73 (5): 120–134.

Porter, M. E. and C. van der Linde. 1995b. Toward a new conception of the environment–competitiveness relationship. *Journal of Economic Perspectives*, 9 (4): 97–118.

Posner, B. Z. and W. H. Schmidt. 1992. Values and the American manager: An update updated. *California Management Review*, 34(3): 80–94.

Preston, L. E. and D. P. O'Bannon. 1997. The corporate social-financial performance relationship. *Business & Society*, 36(4): 419–429.

Ragothaman, S. and D. Carr. 2008. The impact of environmental information disclosures on shareholder returns in a company: An empirical study. *International Journal of Management*, 25(4): 613–620.

Richard, P. J., T. M. Devinney, G. S. Yip, and G. Johnson. 2009. Measuring organizational performance: Towards methodological best practice. *Journal of Management*, 35(3): 718–804.

Roome, N. 1992. Developing environmental management strategies. *Business Strategy and the Environment*, 1(1): 11–24.

Rosenthal, R. 1979. The ''file drawer problem'' and tolerance for null results. *Psychological Bulletin*, 86(3): 638–641.

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

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

Scargle, J. D. 1999. Publication Bias: The 'File-Drawer Problem' in Scientific Inference. *Journal of Scientific Exploration*, 14: 91–106.

Sen, P., M. Roy, and P. Pal. 2015. Exploring role of environmental proactivity in financial performance of manufacturing enterprises: a structural modelling approach. *Journal of Cleaner Production*, 108: 583–594.

Sharfman, M. P. and C. S. Fernando. 2008. Environmental risk management and the cost of capital. *Strategic Management Journal*, *29*(6): 569–592.

Shetty, Y. K. 1979. New Look at Corporate Goals. *California Management Review*, 22(2): 71–79.

Sotorrío, L. L. and J. L. F. Sánchez. 2008. Corporate social responsibility of the most highly reputed European and North American firms. *Journal of Business Ethics*, 82(2): 379–390.

Stanley, T. D. 2001. Wheat from chaff: Meta-analysis as quantitative literature review. *The Journal of Economic Perspectives*, 15(3): 131–150.

Stanley, T. D. 2004. Does unemployment hysteresis falsify the natural rate hypothesis? a meta-regression analysis. *Journal of Economic Surveys*, 18(4): 589–612.

Stanley, T. D. 2008. Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection. *Oxford Bulletin of Economics and Statistics*, 70(1): 103–127.

Stanley, T. D. and H. Doucouliagos. 2010. Picture this: a simple graph that reveals much ado about research. *Journal of Economic Surveys*, 24(1): 170–191.

Stanley, T. D. and H. Doucouliagos. 2012. *Meta-regression analysis in economics and business.* New York: Routledge.

Stanley, T. D. and H. Doucouliagos. 2015. *Neither Fixed nor Random: Weighted Least Squares Meta-Analysis. Statistics in Medicine,* 34 (13): 2116–2127.

Stanley, T. D. and H. Doucouliagos, M. Giles, J. H. Heckemeyer, R. J. Johnston, P. Laroche, J. P. Nelson, M. Paldam, J. Poot, and G. Pugh. 2013. Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*, 27(2): 390–394.

Stanley, T. D., H. Doucouliagos, and S. B. Jarrell. 2008. Meta-regression analysis as the socio-economics of economics research. *The Journal of Socio-Economics*, 37(1): 276–292.

Stanley, T. D. and S. B. Jarrell. 1989. Meta-Regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2): 161–170.

Stanley, T. D. and S. B. Jarrell. 1998. Gender wage discrimination bias? A meta-regression analysis. *Journal of Human Resources*, 33: 947–973.

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

Sutton, A. J. and J. Higgins. 2008. Recent developments in meta-analysis. *Statistics in Medicine*, 27(5): 625–650.

Tatsuo, K. 2010. An analysis of the eco-efficiency and economic performance of Japanese companies. *Asian Business & Management*, 9(2): 209–222.

The Washington Post. 2014. www.washingtonpost.com/blogs/govbeat/wp/2014/09/03/watch-the-u-s-transition-from-a-manufacturing-economy-to-a-service-economy-in-one-gif/. Accessed 29 September 2016.

Thompson, S. G. and S. J. Sharp. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, 18(20): 2693–2708.

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

Tucker, B. P., L. D. Parker, and K. A. Merchant. 2016. With a little help from our friends: An empirical investigation of co-authoring in accounting research. *The British Accounting Review*, 48(2), 185–205.

*Tushman, M. L. and P. Anderson. 1986. Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 31(3): 439–465.

Valentine, J. C., T. D. Pigott, and H. R. Rothstein. 2010. How many studies do you need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(2): 215–247.

Valickova, P., T. Havránek, and R. Horvath. 2015. Financial Development and Economic Growth: A Meta-Analysis. *Journal of Economic Surveys*, 29(3): 506-526.

Villiers, C. de, V. Naiker, and C. J. van Staden. 2011. The effect of board characteristics on firm environmental performance. *Journal of Management*, 37: 1636–1663.

Vishwanathan, P. 2010. The elusive relationship between CSP and CFP: Analyzing four decades of misguided evidence. Paper presented at the Annual Meeting of the Academy of Management, 6–10 August, Quebec.

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

Wagner, M. 2005. How to reconcile environmental and economic performance to improve corporate sustainability: corporate environmental strategies in the European paper industry. *Journal of Environmental Management*, 76(2): 105–118.

Wagner, M. 2008. Empirical influence of environmental management on innovation: Evidence from Europe. *Ecological Economics*, 66: 392–402.

Wagner, M. and S. Schaltegger. 2003. Introduction: How Does Sustainability Performance Relate to Business Competitiveness? *Greener Management International*, 2003(44): 5–16.

Wagner, M., S. Schaltegger, and W. Wehrmeyer. 2001. The relationship between the environmental and economic performance of firms. *Greener Management International*, 34: 94–111.

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

Wernerfelt, B. 1984. A resource-based view of the firm. *Strategic Management Journal*, 5(2): 171–180.

Weidenbaum, M. and S. Vogt. 1987. Takeovers and Stockholders: Winners and Losers. *California Management Review*, 29(4): 157–168.

Xie, S. and K. Hayase. 2007. Corporate environmental performance evaluation: A measurement model and a new concept. *Business Strategy and the Environment*, 16: 148–168.

Yamaguchi, J. and G. C. van Kooten. 2008. Do higher financial returns lead to better environmental performance in North America's forest products sector? *Canadian Journal of Forest Research*, 38(9): 2515–2525.

Yang, M. G. M., P. Hong, and S. B. Modi. 2011. Impact of lean manufacturing and environmental management on business performance: An empirical study of manufacturing firms. *International Journal of Production Economics*, 129(2): 251–261.

Zhu, Q., J. Sarkis, and K. H. Lai. 2007. Initiatives and outcomes of green supply chain management implementation by Chinese manufacturers. *Journal of Environmental Management*, 85(1): 179–189.

Endnotes

[1] There is no clear consensus about the minimum number of studies to be included in a meta-analysis. However, we want to point out that while a sample of two or more studies as suggested, for example, by Valentine et al. (2010) could be used for the calculation of weighted mean effects meta-analysis, meta-regression analysis with a broad set of explanatory variables requires a sufficient set of observed effect sizes to reach reliable regression estimates.

[2] Already Endrikat et al. (2014, 736) notice that although "Horváthová (2010) labeled her study as a meta-analysis, she obviously applied a vote count procedure."

[3] In theoretical literature CEP is not treated separately from the broader term corporate social performance (CSP) including CEP, although environmental performance has a biophysical nature (Salzmann et al. 2005). The theory of corporate social performance can be transferred to CEP as literature states the validity of the underlying hypotheses for both artefacts (Endrikat et al. 2014).

[4] A detailed discussion of the theories described in Table 2 is available from Appendix A or from one of the existing reviews on the CEP-CFP theories (see, among others, Guenther and Hoppe 2014).

[5] Due to limited access we could not search the E-Journals and Environment Complete database. However, in our opinion both databases cover largely the same publication outlets as the other six databases used in this study.

[6] The search command consists of keywords linked by Boolean operators with the following logic: (i) terms relating to CEP ("social*", "environment*", "green", "pollut*", "ISO", "emission", "emit*"), (ii) terms relating to CFP ("firm", "compan*", "business", "corporat*", "financ*", "economic*", "value", "performance", "pay", "market", "return"), and (iii) terms relating to the applied methodology ("empirical", "statistical", "test", "analy*", "relation*", "survey*"). Compared to Endrikat et al. (2014) we extended the search command by three further keywords for CEP ("ISO", "emission", "emit*").

[7] As the majority of studies use firm-level data, we exclude portfolio study results from our sample.

[8] Moreover, perceptual measures are not included in the analysis due to the restricted comparability to quantitative data used in the majority of primary studies. As the data availability does not allow the analysis of a further explanatory variable controlling this difference, we had to exclude the corresponding studies. For a further description of the included CEP and CFP measures, please refer to the definitions of the explanatory variables in section 3 and Appendix C.

[9] Appendix F shows all studies with the number of Google Scholar citations, the number of extracted effect sizes, the mean number of firm/firm year observations, the examined industries and countries separately for each study. Appendix G presents the publication outlets covered by the studies in our sample. Appendix H gives an overview of the publication years covered by the sample.

[10] For the calculation of the effect sizes please refer to Appendix I.

[11] Concerning the operationalization of CEP in the measurement of the CEP-CFP relation in primary studies, the direction of the influence of CEP on the effect size is not unique (Albertini 2013). For example, CEP sometimes refers to the total amount of waste (negative) that has to be treated differently from the amount of reduced emissions (positive). Hence, the sign of the effect size is unified across studies so that environmental consciousness always leads to higher CEP.

[12] The existence of substantial heterogeneity among the integrated N effect sizes is tested by Cochran's Q-test (Cochran 1954; Higgins and Thompson 2002) and the I^2 statistics, which shows the proportion of variation that is due to heterogeneity rather than sampling error (Higgins et al. 2003).

[13] Alternative approaches for the identification of publication selection bias are Rosenthal's failsafe N (Rosenthal 1979) as used by Endrikat et al. (2014) or Hedges' maximum likelihood publication selection estimator (Hedges 1992). However, both methods show a number of severe problems (Begg and Berlin 1988; Doucouliagos and Stanley 2013; Scargle 1999) and thus the MRA approach to model publication selection is widely been accepted as the superior procedure (Stanley 2001).

[14] The meta-regression error variance is assumed to be known rather than estimated.

[15] For the detailed results of the heterogeneity test and all further univariate calculations, please refer to Appendix J.

[16] In the random effects case, the effect sizes are weighted by the inverse sum of the standard error SE(r_i) and the between-study variance τ^2.

[17] For the funnel plot, please refer to Appendix K.

[18] For detailed results of the funnel asymmetry test please refer to Appendix J.

[19] The reference group for this variable captures countries with a mixed legal system.

[20] Due to the high collinearity between GDP and GDP², we integrate the two variables separately in the estimation models.

[21] The corresponding calculation according to the parsimonious model (1) from Table 4 yields the following value for the CEP-CFP relation:

0.427=0.273+0*(-0.366)

+ 1*0.274+0*(-0.027)+0*(-0.218)+0*(-0.157)+1*0.109+0*0.086 (H1)

+ 1*0.022+0*(-0.069) (H2)

+ 0*(-0.042)+0*0.035+0*(-0.125)+0*0.016 (H3)

+ 1*(-0.298)+4.361*0.003+282.047*(-0.000)+0*(-0.240)+0*(-0.172)

+ 1*0.111+0.143*0.125+2.207*(-0.010)+0.487*0.025+2.262*0.040

+ 0*0.030+1*(-0.017)+1*(-0.024)+0*0.024+0*0.015+1*(-0.085)

+ 1*(-0.050) (H4)

[22] Based on a review of more than 22,000 partial correlations reported in previous meta-analyses, Doucouliagos (2011) classifies partial correlations between 0.07 and 0.17 as ‘small’, effects above 0.17 and below 0.33 as ‘moderate’, and values above 0.33 as ‘large’.

[23] This list includes all references cited in the text. A full list of the studies included in the meta-analysis sample is available from Appendix E.