ENVIRONMENTAL RESEARCH
LETTERS

TOPICAL REVIEW • OPEN ACCESS

A meta-analysis of the relationship between companies' greenhouse gas emissions and financial performance

To cite this article: Jan Taeke Galama and Bert Scholtens 2021 *Environ. Res. Lett.* **16** 043006

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
24 February 2020

REVISED
15 January 2021

ACCEPTED FOR PUBLICATION
22 January 2021

PUBLISHED
25 March 2021

TOPICAL REVIEW

A meta-analysis of the relationship between companies' greenhouse gas emissions and financial performance

Jan Taeke Galama[1] and Bert Scholtens[1,2]

[1] Department of Economics, Econometrics and Finance, University of Groningen, Groningen, The Netherlands
[2] School of Management, University of St Andrews, Saint Andrews, Scotland, United Kingdom

E-mail: l.j.r.scholtens@rug.nl

Keywords: meta-analysis, firms, greenhouse gas emissions, financial performance, research design, industry effects, policy stringency

Abstract

We study how the business and economics literature investigates how companies' greenhouse gas (GHG) emissions relate to their financial performance. To this extent, we undertake a meta-analysis to help us gauge the role of using highly different constructs and measurement techniques employed in this literature. Our study includes 74 effect sizes from 34 studies, covering 107 605 observations for the period 1997–2019. We establish a significant association between corporate GHG emissions and financial performance. It shows that companies with lower emissions have better financial performance. We find that the type of emission or financial performance indicator is not significant. The industry to which the firms in the sample studies belong does seems to matter slightly. We further establish that the relationship between GHG emissions and financial performance is especially pronounced for firms operating in countries with the most stringent carbon policies.

1. Introduction

It is well established that there is a relationship between firms' environmental performance (hereafter: CEP) and their financial performance (hereafter: CFP). Some studies show that corporate environmental performance (CEP) and corporate financial performance (CFP) are negatively associated (e.g. Hassel *et al* 2005, Qi *et al* 2014, Misani and Pogutz 2015, Brouwers *et al* 2018), whereas others find a positive relationship (e.g. Hart and Ahuja 1996, Russo and Fouts 1997, King and Lenox 2001, Wang *et al* 2014, Makridou *et al* 2019). There are also studies which arrive at a neutral effect (e.g. Waddock and Graves 1997, Konar and Cohen 2001). The diverging findings seem to result from the wide range of methods employed and the variety in indicators used to measure both CEP and CFP (Guenther *et al* 2011, Albertini 2013, Dam and Scholtens 2015), as well as from moderating factors like industry and country characteristics (Albertini 2013, Dixon-Fowler *et al* 2013, Endrikat *et al* 2014).

We concentrate on how firms greenhouse gas (GHG) emissions relate to their financial performance and investigate factors that might influence the GHG–CFP relationship. We focus on GHG emissions as their reduction is crucial to achieve the objectives of the Paris 2015 agreement in relation to mitigating climate change (Fujii *et al* 2013, Trinks *et al* 2018). We employ meta-analysis of studies after the relationship between corporate GHG performance and CFP to summarize, evaluate, and analyze empirical findings in this research field (Kirca and Yaprak 2010). Since the majority of studies included in our review do not measure the direction of the causality of the relationship, we cannot relate to this in our study as we are confined to the nature and scope of the studies included (Hunter *et al* 1982).

In this study, we first investigate the overall relationship between firms' GHG emissions and financial performance. Then, we study whether the type of reporting (voluntary, mandatory) plays a role. Third is that we compare the impact of absolute measures with those of relative ones. Fourth, we compare accounting-based measures of financial performance with financial market based ones. We also investigate whether industry affiliation matters, in particular the generic GHG emission intensity per industry, in relation to the association between firms' GHG and financial performance. Lastly, we study the effect of climate policy stringency.

2. Background and hypotheses

Numerous studies relate social and environmental performance to financial performance: Friede *et al* (2015) report there are more than 2000 of such studies. Then, meta-analysis is useful as it provides an integrated perspective of the results from using various data sources, control variables and estimation techniques (see table 1 for an overview).

Meta-analysis by Orlitzky *et al* (2003) and Alloche and Laroche (2005) documents a significant and positive relationship between companies' social and environmental performance and their financial performance. However, they also observe that the research design employed significantly influences this relationship. Social and environmental performance (i.e. corporate social performance; hereafter CSP) is of a very broad and diffuse nature, making it hard to provide a sound comparison and analysis. Dixon-Fowler *et al* (2013) try to bring focus and perform a meta-study on the relationship between environmental and financial performance. Here too, it shows there is a significant and positive association. They find that the association is significantly weaker when CEP is measured by emissions compared to other environmental performance measures. They report that contingencies (e.g. differences in firms' size) and methodological issues (e.g. mandatory versus voluntary reporting) moderate the CEP–CFP relationship. Vishwanathan *et al* (2020) concentrate on the transmission mechanisms between CSP and CFP. They establish that CSP influences financial performance via firm reputation, stakeholder reciprocation, firm risk, and innovation capacity. Albertini (2013) finds that the CEP–CFP relationship is influenced by the constructs used for both environmental and financial performance, regional differences, industry, and the period studied. Endrikat *et al* (2014) investigate both the direction of the causality and the multidimensionality of constructs. They find a positive relationship between CEP and CFP, and this appears to be partially bidirectional. Lewandowski (2015) and Busch and Lewandowski (2018) relate a firm's total carbon dioxide emissions to its financial performance and arrive at an inverse relationship between the two.

Climate change mitigation and adaptation receives increasing attention as governments, consumers, and financial market participants are increasingly concerned about global warming (Wang *et al* 2014, Trinks *et al* 2018). Carbon regulation has emerged in several countries and regions and emissions have become a cost factor for business (Clarkson *et al* 2015, Trinks *et al* 2020).

Our meta-analysis aims to contribute to this literature from a range of perspectives: It studies the post-Kyoto era, focuses on the corporate level, uses GHG emissions as a CEP measure, relies on a systematic selection and analysis of the sample studies, accounts for industry affiliation, and investigates whether climate policy stringency is a vector. We regard the Kyoto Protocol as a breakpoint in climate policy as it contains the possibility for internationally legally binding emission targets for industrialized countries that trickled down into targets for business (Böhringer 2003). Therefore, we concentrate on studies using sample periods from 1997 onwards. Next, we take the corporate perspective, as it is primarily businesses who emit GHGs in the production and distribution process (see World Bank 2019). Of the existing meta-studies, only Busch and Lewandowski (2018) explicitly focus on the relationship between firms' GHG emissions and their financial performance. However, they do not seem to use a particular algorithm to select studies and their sample cannot be replicated. We deem this highly important and will select studies based on clear and transparent selection criteria. GHG emissions refer to the amount of carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorocarbons, sulphur hexafluoride, and nitrogen trifluoride emissions (IPCC 2014). We define corporate GHG performance as the inverse of the GHG emissions of firms. As such, low (high) amounts of GHG emissions refer to high (low) GHG performance (see Misani and Pogutz 2015). Studies using GHG emissions as a proxy for CEP collect their data either from mandatory or voluntary reporting schemes. We investigate whether this influences the results. Further, GHG emissions can be in absolute or relative terms. From an economic perspective, it is relative terms that matter. However, from an environmental point of view, it is the absolute amount of emissions that goes into the atmosphere that is relevant. Therefore, we examine if and how scaling influences the results. Further, CFP can be measured using accounting-based indicators or market-based indicators. Some studies argue CEP is more strongly related to (contemporaneous and forward-looking) market-based measures of CFP than to (backward looking) accounting-based indicators of CFP (Dixon-Fowler *et al* 2013). Previous studies have investigated both measures and provide conflicting results. Consequently, we investigate how measuring financial performance relates to GHG performance. Next, we study if industry-specifics play a role. Delmas *et al* (2011) argue that environmental regulations for the most polluting industries are stricter and that polluting firms employ different strategies for reducing their emissions. Our study explicitly accounts for industry affiliation. We investigate whether the relationship between corporate GHG performance and CFP is different for studies after firms in polluting-intense industries compared to studies that do not differentiate in this regard. Lastly, Endrikat *et al* (2014) suggest that including country-specific factors, such as differences in regulatory environmental systems, might play a role. In

Table 1. Summary generic meta-studies and their main findings.

Authors	Relationship	Period	Region	*n*	Main selection criterion	Key findings
Orlitzky *et al* (2003)	CSP–CFP	1972–1997	Global	17	Period	CSP–CFP relationship is positive in nature and several factors moderate the relationship.
Alloche and Laroche (2005)	CSP–CFP	1972–1996	Global	82	Period and CSP construct	CSP is strongly related to CFP and both the measurement and method of the empirical study affect the outcomes.
Dixon-Fowler *et al* (2013)	CEP–CFP	1970–2009	Global	39	Period	CFP is significantly influenced by CEP. Several contingencies moderate the relationship.
Albertini (2013)	Environmental Management—CFP	1975–2011	Global	52	Period	A positive relationship between environmental management and CFP. The relationship is influenced by performance measures and other contingencies.
Endrikat *et al* (2014)	CEP–CFP/ CFP–CEP/ bidirectional	All	Global	149	CEP–CFP measures	The results indicate that there is a positive and partially bidirectional relationship between CEP and CFP. They find several moderating effects.
Busch and Lewandoski (2018)	Carbon Performance	All	Global	32	CEP measure	Corporate carbon performance positively related to CFP. Outcomes vary with CFP measures.
Vishwanathan *et al* (2020)	CSP–CFP	1978–2016	Global	344	CSP construct	CSP enhances firm reputation, increases stakeholder reciprocation, mitigates firm risk, and strengthens innovation capacity

Note: This table is an overview of the existing meta-studies about the relationship between corporate social performance (CSP) or corporate environmental performance (CEP) and corporate financial performance (CFP). It includes information about the studied relationship, the period covered, the investigated region, the number of studies included in the study (*N*), the main sample selection criteria, and the main findings of the study.

this regard, several economists argue that cap-and-trade systems like emission trading systems (ETSs) are the most cost-effective way to reduce the environmental impact of countries (e.g. Bowen 2018). These carbon-pricing mechanisms cover 11 gigatons of CO2 emissions, representing 20% of the global emissions (World Bank 2019). Clarkson *et al* (2015) posit that carbon emissions affect firm valuation only to the extent that a firm's emissions exceed its carbon allowances under a cap and trade system and the extent of its inability to pass on carbon related compliance costs to consumers and end users. Czerny and Letmathe (2017) find that GHG emissions were not reduced cost-effectively. They argue that companies' intrinsic values prevail over economic incentives from the ETS regarding carbon reduction. Both Clarkson *et al* (2015) and Czerny and Letmathe (2017) relate to the European Union's ETS only. To investigate the role of climate policy, we will investigate the impact of the policy stringency of the ETS.

Kuo *et al* (2010) find a positive relationship between GHG and financial performance and attribute this to eco-efficiency. Eco-efficiency implies that productivity gains through reduction of materials use, improvements in the manufacturing processes, and utilization of waste can improve the operational efficiency of firms (Kuo *et al* 2010). Improved efficiency via emission reduction and the utilization of by-products and waste can lead to both lower costs and more innovation, improving firms' comparative advantage (Orsato 2006, Kuo *et al* 2010). Institutional investors may require companies to take their responsibility and become more eco-efficient too (Trinks *et al* 2018, 2020). Consumers may avoid buying products from companies that have poor GHG performance. Then, firms can improve their financial performance by reaping the reputational benefits associated with cleaner production (Hart and Ahuja 1996). When investments in GHG emission reduction require significant up-front investments, costs may outweigh the benefits of the investment and therefore weaken firms' financial performance (Brouwers *et al* 2018). Fujii *et al* (2013) argue that emission reduction may negatively affect a company's competitive position as resources are allocated to non-core business operations. Enkvist *et al* (2007) indicate that the costs of emission reduction can differ widely between specific types of technology and over time. Therefore, we first investigate how GHG performance associates with CFP. In this regard, the following two competing hypotheses are tested:

***H1A**: The association between GHG and financial performance is positive*

***H1B:** The association between GHG and financial performance is negative*

GHG emissions are administered via voluntary or mandatory reporting. Voluntary reporting schemes collect their data mostly by questionnaires and surveys, like the Carbon Disclosure Project. Voluntary reporting might result in a self-selection bias, allows for different methodologies, and usually lacks external verification (Perrault and Clark 2010, Chen and Gao 2012). In contrast, data collected via mandatory reporting is based on formal rules, which allows for comparison between industries and countries and over time (Perrault and Clark 2010). However, even data from mandatory reporting schemes can be biased, for example when firms can select the plants eligible for reporting, emission factors, and the specific way to measure emissions (Sullivan and Gouldson 2012). Several studies find that greater consideration for the impact of corporate activities on the environment and control of GHG emission may help reduce costs (such as waste management, energy and water consumption) and achieve benefits (improve reputation, increase revenues, improve competitiveness) (see Jiang and Bansal 2003). This may encourage firms to voluntarily disclose and reduce their GHG emissions (see Arimura *et al* 2008). However, Bansal and Roth (2000) and Lyon and Maxwell (2011) point out that there might also be greenwashing going on in this regard. Therefore, several jurisdictions opt for mandatory disclosure (such as the Norway, Singapore, UK) and hope that such disclosure will incentivize innovation and environmental performance (see Tang and Demeritt 2018). Of course, raising awareness in this way too may impact the corporate activities and environmental performance, but there is less scope for greenwashing. Notwithstanding, especially companies in energy intense industries will already have had emissions on their radar, but this might not have been the case elsewhere. Then, mandatory reporting may have resulted in the realization of new areas to manage costs and benefits in the latter industries. However, as the role of energy will have been only minor in the industries that were not already focused on emissions, one may not expect a substantial impact on the relationship between emissions and financial performance. Therefore, it is not likely this relationship will be stronger in the case of mandatory than with voluntary reporting (Tang and Demeritt 2018). In all, we think it is not possible to postulate whether the relationship between GHG emissions and financial performance is stronger in either of the two regimes.

Thus, we assume it is not evident which type of reporting more closely relates to CFP. Hence, we test the following hypothesis:

***H2:** The type of reporting scheme influences the results in GHG and financial performance studies*

Studies measure GHG emissions with either absolute or relative indicators (Slawinski *et al* 2017). Absolute emissions reflect the physical emissions of a firm in a given period of time. Relative emissions relate these emissions to firms' key characteristics (e.g. number of employees, sales, revenues, costs), commonly labelled as carbon intensity or efficiency (Kuik and Mulder 2004, see Trinks *et al* 2020, for a critical reflection). We want to stress that in this regard the sample studies are not always clear what exactly is being used as the denominator in relation to the emissions, implying that the literature is subject to the homogeneity problem. Clarkson *et al* (2015) argue that absolute emissions have to be used to determine the costs of businesses as the acquisition of emission rights is based on the firms' overall emissions. Absolute emissions of businesses directly inform about their contribution to climate change (Ekwurzel *et al* 2017). GHG performance measured by absolute indicators should therefore be more strongly related to CFP. In contrast, Olsthoorn *et al* (2001) argue that emissions of firms have to be judged relative to their peers to allow for comparison. This is because financial market participants incorporate the extent to which the business model relates to GHG emissions and they compare different prospects (Trinks *et al* 2018). As such, relative GHG performance would be more strongly related to CFP than absolute GHG. Therefore, we study whether the nature of the measure for GHG performance influences the relationship with CFP and test the following two competing hypotheses:

***H3A:** GHG performance influences financial performance more strongly when it is measured by relative than by absolute emissions*

***H3B:** GHG performance influences financial performance more strongly when it is measured by absolute than by relative emissions*

Further, several measures are used to proxy for CFP. Most studies use either accounting-based or market-based measures (Albertini 2013), but sometimes reputation, stakeholder reciprocation, firm risk, and innovation capacity is used too (Vishwanathan *et al* 2020). Accounting-based measures usually encompass indicators like return on assets (ROAs), return on equity (ROE), or return on sales (ROS) (Danso *et al* 2019). These indicators reflect the internal capabilities of the firm to generate value, rather than external perceptions of performance (Orlitzky *et al* 2003). They are of a backward-looking nature as the information about the constituting elements is available with some time lag. In contrast, market-based measures are of a more contemporaneous nature and also include market expectations about future conduct and performance (Dam and Scholtens 2015). Examples are (excess) stock market returns, stock return volatility, price-earnings ratio, price per share, and earnings per share (Dowell *et al* 2000, Orlitzky *et al* 2003). Albertini (2013) and Orlitzky *et al* (2003) find that accounting-based indicators are more closely related to CEP than market-based ones. Ambec and Lanoie (2008) reason that investments in GHG performance will be converted into better future accounting-based performance (Ambec and Lanoie 2008). In contrast, Dixon-Fowler *et al* (2013) find that CEP more closely relates to market-based performance. This would suggest that investors value carbon emissions and use off-balance sheet valuation discounts for GHG emission (Griffin *et al* 2017). This might be the case if outstanding GHG performance reduces regulatory risk and can become of increasing value in the case of future changes in carbon regulation (Albertini 2013). Therefore, we test:

H4A: *Corporate environmental performance is more strongly related to prior market-based than to prior accounting-based financial performance*

H4B: *Corporate environmental performance is more strongly related to prior accounting-based than to prior market-based financial performance*

The relationship between CEP and CFP can differ due to different combinations of production factor inputs and technology usage (Konar and Cohen 2001). Such combinations vary between firms and per industry. Hart and Ahuja (1996) find that the largest impacts on CFP accrue to 'high polluters' since they can make low-cost improvements; in less-polluting industries, investments in CEP tend to become increasingly expensive. Delmas *et al* (2011) find that this changes over time as additional emission reduction becomes increasingly more costly. So far, the focus in CEP–CFP studies has primarily been on industrial companies, as these are the ones concerned most with toxic and hazardous emissions (King and Lenox 2001). Some studies concentrate on particular subsectors (Van der Goot and Scholtens 2015) and find clear differences between these. Others rely on industry-wide data to arrive at generalizable results (Albertini 2013). Most of these studies suggest that the GHG intensity of the industry in which a company operates affects the results. Therefore, we test the following hypothesis:

H5: *The relationship between GHG performance and CFP is strongest in the most polluting sectors.*

An ETS puts a price on GHG emissions. In general, these systems consist of tradable emission permits and an overall cap on emission that decreases over time (Alkhurst *et al* 2003, Van der Goot and Scholtens 2015). An ETS leaves companies with three alternative strategies: reducing GHG emissions to meet the requirements, buy emission rights, or reduce emissions to a level below the legal requirements and sell the excess emission rights (Sandoff and Schaad 2009). Since all strategies affect the costs of emissions, Policy stringency will influence the relationship between GHG performance and CFP (Czerny and Letmathe 2017). Stringency particularly relates to the proportion of GHG in the jurisdiction covered, the number of industries participating, the price of emission rights, and the amount of emission allowances distributed under free allocation or auctioning (World Bank 2019). Firms participating in ETSs that are more stringent face more carbon constraints (Joltreau and Sommerfeld 2018). A relative stringent policy imposes more costs on firms, as they have to invest more than firms under less stringent ones. A stringent policy also increases the monitoring and reporting costs of firms. Deschenes (2018) argues that a more stringent policy leads to worse financial performance and competitiveness compared to firms operating under less stringent regimes. Next to the impact on the firm, it is important to realize that ETSs allocate the costs of externalities that are otherwise fully borne by society. We hypothesize that the relationship between GHG and CFP may be more stronger (more positive) in jurisdictions with more stringent climate policy regimes.

H6: *The relationship between GHG performance and CFP is stronger for firms operating in countries with more stringent climate policy than for firms in countries with weak policy stringency.*

3. Methodology

To test our hypotheses, we use meta-analysis to investigate the empirical findings regarding the relationship between GHG and CFP. Results from a meta-analysis may include a more precise estimate of the effect of a construct than any individual study contributing to the pooled analysis (Tavakol 2018). First, we present the way in which we sample studies. Then, we describe the effect sizes and coding procedures. Thirdly, we reflect on the meta-analytical procedure.

3.1. Sampling

In a meta-analysis, the literature included has to be systematically selected (Stanley and Doucouliagos 2012). In this regard, we rely on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) method, which consists of four stages in data collection: identification, screening, eligibility, and inclusion (Moher *et al* 2010). To incorporate all relevant studies, an extensive search with a broad set of keywords was conducted. We used the following search (combinations): *corporate environmental performance, CEP, environmental performance, corporate financial performance, financial performance, CFP, does it pay to be green, when does it pay to be green, carbon performance, GHG performance, climate change, GHG emissions, CO2 emissions, environmental management, environmental regulation,* and *carbon-pricing*. The (electronic) search was conducted using EBSCO, ScienceDirect, JSTOR, Emerald, and Google Scholar, and we selected peer-reviewed studies. In contrast, other meta-analyses (e.g. Dixon-Fowler *et al* 2013, Endrikat *et al* 2014, Busch and Lewandoski, 2018) also include papers based on a search in references of non-academic papers (i.e. not being peer-reviewed) and as well as conference presentations. As this might lead to systematic bias (Hunter *et al* 1982), we do not employ these. We limit the study to peer-reviewed academic work; we also refrain from including our own studies in the meta-analysis.

Our search based on keywords yielded an initial sample of 73 studies. Next, we implemented four inclusion criteria. First, we include only studies on GHG–CFP that rely on data from 1997 onwards. This is because of the Kyoto Protocol which marks the start of a new era of climate policy (Böhringer 2003). Because of the resulting shift in perception of the stakeholders towards impact of climate change policy, papers including data from before 1997 might yield different results compared to more recent studies (see also Endrikat *et al* 2014). Second, since we are interested in the effect sizes regarding GHG emission, we only include studies that measure the relationship between GHG emissions and CFP. We point out that the sample studies may use different measures in this regard. Most GHG emissions are measured by CO_2e scope two emissions but more than half of the studies does not disclose in a transparent manner. This is a problem in most of the literature, where business and economics scholars use metrics they are not very familiar with. However, the same lack of transparency occurs with financial performance, especially accounting performance. Financial performance is measured via accounting and market data and we investigate whether the findings differ in case either of the two is used. We also point out that the potential of the multiplicity of data in the sample studies may lead to variability in the results of the meta-analysis (Tendal *et al* 2011). In fact, most studies do not include a detailed account of the sampling procedure regarding the selection of countries, industries, and firms or the period studied. This is problematic and requires disciplining in this regards within the field of business and economics as it does not allow for full replication of the results. Third, to allow for comparison, the studies have to report sample sizes and correlation coefficients or statistics that can be converted into these. Finally, we only include results from continuous variable studies as it is in general not possible to compare results from binary regressions (e.g. probit and logit studies) (see Hunter *et al* 1982). Likewise, we exclude event studies as their methodology is highly different from that of other estimates (Stanley and Doucouliagos 2012). As a result, and reported in table 2, our final sample consists of 34 articles.

3.2. Coding

The effect sizes of the individual studies are the main unit of our analysis. Effect sizes are gathered from two types of statistics: Pearson product-moment correlations and partial-correlations. Pearson product-movement are derived from the correlation table in the empirical studies. For studies that did not report correlation tables, the effect sizes (r) are calculated from the reported t-statistics and the degrees of freedom; for studies that do not report the t-statistic, it is calculated backward from the standard errors, significance level, or probability values. Studies often report more than one relationship because they use multiple constructs (Albertini 2013). Then, two approaches can be used to deal with multiple measures from independent studies, namely treating them as independent effect sizes or representing each study by a single effect size. Using a single observation for each primary study leads to loss of information, as averaging has to take place. Therefore, we include all observations from reported CFP constructs (e.g. Tobin's Q, ROA, ROE) and from GHG performance constructs (e.g. absolute, relative). In line with Stanley and Doucouliagos (2012), the result from the model with the highest adjusted R-squared is included. Accordingly, from our 34 studies a total of 74 effect sizes are extracted ($k = 74$), with 107 605 observations ($n = 107\,605$). Appendix A provides an overview of these 74 effect sizes and the corresponding sample size. The key features of our sample are depicted in figure 1. Panel A in figure 1 summarizes and shows the majority of effect sizes are positive: 52 effect sizes indicate a positive relationship; 21 effect sizes indicate a negative relationship, and one effect size does not show any significant relationship between the constructs. Most effect sizes are extracted from Pearson product-movement reporting in the correlation table, others were calculated from the t-statistic and the degrees of freedom. Panel B shows that the sample studies included both market and accounting-based indicators to measure CFP: 47 observations are based on accounting-based indicators and 27 on market-based indicators of CFP. For both, the majority of the effect sizes are positive. Panel C of figure 1 displays the characteristics of the GHG performance construct used in the 34 studies. Most studies use relative emissions for CEP and the majority of observations are collected based on voluntary reporting schemes. Panel D provides the industry composition and shows that firms from the

Table 2. Key characteristics of the individual sample studies.

					CFP measure		CEP measure	
Paper	Authors	Period	Region/Country	Industries	Accounting-based	Market-based	Indicator specification	Reporting scheme
1	Aggarwal and Dow (2012)	2008–2009	US	Multiple	ROA	Tobin's *Q*	Relative	Voluntary
2	Brouwers *et al* (2018)	2005–2012	Europe	Multiple	ROA, ROE	Tobin's *Q*	Relative	Mandatory
3	Brzobahaty and Jansky (2010)	2004–2006	Czech Republic	Multiple polluting	Inpa, Inra, Inca		Relative	Mandatory
4	Busch and Hoffmann (2011)	2005–2007	Global	Multiple	ROA, ROE	Tobin's *Q*	Relative	Voluntary
5	Busch, Lehmann, Hoffmann (2012)	2003–2009	Global	Multiple	D/a, CF To Assets	Total market risk, unsystematic risk, systematic risk	Absolute	Voluntary
6	Chakrabarty and Wang (2013)	2001–2009	US	Manufacturing, utilities	SALES effectiveness, ROE, ROA		Absolute	Mandatory
7	Chapple *et al* (2013)	2007–2009	Australia	Multiple	V, Earnings		Relative	Mandatory
8	Clarkson *et al* (2015)	2006–2009	Europe		ROA		Absolute	Mandatory
9	Clarkson *et al* (2015)	2005–2006	Australia	Multiple	ROA	Tobin's *Q*	Absolute	Mandatory
10	Dangelico and Pontrandolfo (2015)	2011	Italy	Multiple	Market Performance		Relative	Mandatory
11	Delmas, Nairn-Birch and Lim (2015)	2004–2008	US	Multiple	ROA, ROE	Tobin's *Q*	Absolute	Voluntary
12	Fujii *et al* (2013)	2006–2008	Japan	Manufacturing	ROA, CT, ROS		Relative	Voluntary
13	Gallego-Alvarez, Garcia-Sanchez, and de Silva Vieira (2014)	2006–2009	Global		ROA		Relative	Voluntary
14	Gallego-Álvarez, Segura, and Martínez-Ferrero (2015)	2006–2009	Global	Multiple	ROA, ROE		Absolute	Voluntary
15	Griffin *et al* (2017)	2006–2012	Global	Multiple		PRCC	Relative	Voluntary
16	Hatakeda, Kokubu, Kajiwara, Nishitani (2012)	2006–2008	Japan	Manufacturing	Profitability	Tobin's *Q*	Relative	Mandatory
17	Kuo, Huang, Wu (2010)	2001–2006	Japan	Chemical automobile, electronic	Net income		Absolute	Voluntary
18	Iwata and Okada (2011)	2004–2008	Japan	Manufacturing	ROA, ROE, ROI, ROIC, ROS	Tobin's *Q*	Relative	Voluntary

(Continued)

Table 2. (Continued.)

					CFP measure		CEP measure	
Paper	Authors	Period	Region/Country	Industries	Accounting-based	Market-based	Indicator specification	Reporting scheme
19	Jung *et al* (2016)	2009–2013	Australia	Multiple	Cost of debt		Relative	Mandatory
20	Kim *et al* (2015)	2007–2011	South Korea	Multiple	COE, ROA		Relative	Mandatory
21	Lannelongue *et al* (2015)	2011	Spain	Multiple	ROA, ROE, Profits		Absolute	Voluntary
22	Lee and Min (2015)	2001–2010	Japan	Manufacturing		Tobin's *Q*	Relative	Voluntary
23	Luo and Tang (2014)	2011	Australia	Multiple	Direct exposure to tax	Market return	Relative	Mandatory
24	Makridou *et al* (2019)	2006–2014	Europe	Multiple	EBITDA, current ratio, solvency ratio		Relative	Voluntary
25	Matsumura *et al* (2014)	2006–2008	US	Multiple		MKTE	Relative	Voluntary
26	Misani and Pogutz (2015)	2007–2013	Global	Industrial	ROS, ROA	Tobin's *Q*	Relative	Voluntary
27	Nishitani and Kokubu (2012)	2006–2008	Japan	Manufacturing		Tobin's *Q*	Relative	Mandatory
28	Rokhmawati *et al* (2015)	2015	Indonesia	Manufacturing	ROA		Relative	Mandatory
29	Saka and Oshika (2014)	2012	Japan	Multiple		MVE	Absolute	Voluntary
30	Secinaro. Brescia, Calandra and Saiti (2020)	2013–2017	Europe	Multiple	ROE		Relative	Mandatory
31	Tatsuo (2010)	2003	Japan	Manufacturing	ROA		Relative	Mandatory
32	Trumpp and Guenther (2017)	2008–2012	Global	Multiple	ROA, TSR		Relative	Voluntary
33	Wang *et al* (2014)	2010	Australia	Multiple	Sales	Tobin's *Q*	Absolute	Voluntary
34	Qi *et al* (2014)	1999–2010	China	Multiple	ROA		Relative	Voluntary

Note: Table 2 presents the included studies. It gives information about the period and region covered by the study, the industry of interest, and information about the CEP and corporate GHG performance indicators. A total of 34 studies from the period 1997–2019 are included in the sample. ROA = Return on Assets; ROE = return on equity; ROS = return on sales; ROI = return on investment; ROIC = return on invested capital; Inpa = Profit over assets; Lnra = revenues over assets; Inca = cost over assets; PRCC = stock price three months after fiscal year-end CF = cashflow; EBITDA = Earnings before interest, taxes, deprecation, and amortization; V = market value of common equity; CT = capital turnover CoE = cost of equity; MKTE = Market value total equity MVE = market value equity; TSR = total stock return manufacturing industry make up two fifths of the sample firms.

To come to grips with policy stringency, we use the Climate Change Performance Index (hereafter: CCPI; Burck *et al* 2016), the Climate Change Cooperation Index (hereafter: C3-I; Bernauer and Böhmelt 2013), and the Climate Action Tracker (hereafter CAT; see https://climateactiontracker.org/countries/). Details about these indicators are in Table D1 in the Appendix.

The CCPI tracks efforts of countries to address climate change. It covers 58 countries between 2005 and 2019. C3-I offers a dataset including 172 countries for the period 1996–2008. Both indices capture overall performance scores as well as performance in terms of political behavior and emissions. The methodologies are closely related; they evaluate the emission component based on trends and emission levels. The policy component is assessed by expert assessment in CCPI but based on observed behavior in C3-I. Both measure historical output and emission trends in a wider range of environmental policies and do not measure the future carbon constraints faced by companies (Bernauer and Böhmelt 2013). As their methodologies are slightly different and the indices do not fully cover the whole period of our study, we proceed as follows: The CCPI was extracted from the website accompanying Burck *et al* (2016) - this data is available from 2005 onwards and we used the 2016 data; the codebook and data for C3-I were provided by Böhmelt (2013). Reassuringly though, for overlapping years, it shows that both indices yield identical country ranking. Therefore, we use CCPI as our basis for ranking countries for the periods 1997–2008 and C3-I for 2009–2019. To separate studies based on ETS stringency in the range of countries included, we construct 'study ranks' with the help of the country ranks. For studies conducted in a particular year in a specific country, this rank relates to the median rank of the country ranks of the year before the study, the year after the study, and the study year. By averaging over a three-year period, we reduce the effect of one-off events, like novel policy intentions of governments. Such events may initially improve the country score, but may not always persist (see Burck *et al* 2016). For studies that collect their data in a single country over multiple years, we use the average median rank of the country over this period. For studies with multiple countries over multiple years, the average median rank of the countries is weighted by the number of observations per country. The use of study ranks allows us to assess climate policy stringency of the sample countries in each study, and compare with other studies (Botta and Kozluk 2014). To this extent, we differentiate along four groups of studies according to the climate policy stringency of their sample. When studies do not provide information about the number of observations from individual countries, they are excluded from the ranking. This approach allows dividing studies into four groups with the use of the two indexes, even though the scales and methodologies of both indices are not exactly the same.

In contrast to these two indices, Climate Action Tracker (CAT) assesses and ranks the intentions and progress of governments towards reaching the globally agreed aim of holding global warming below 2 °C. Hence, this is a more contemporaneous and forward looking assessment of stringency. CAT scores are based on the effect of current policies on emissions, the impact of pledges and targets, and fair share and comparability of effort. CAT ranks countries on a scale from critically insufficient to role models (New Climate; Climate Analytics 2011). Further, it accounts for regional effects, assuming that ETS stringency in a particular region will be higher when both individual reduction targets and actions of countries related to achieve the Paris Agreements are more ambitious. Hence, CAT provides a more contemporaneous and forward looking perspective. Studies are grouped based on CAT evaluation of the region in which they are performed: sufficient, medium, moderate, and insufficient (due to small subsamples, we combine medium and moderate). Appendix D1 highlights the key features of the three stringency indices used in this study. Appendix D2 relates the studies to the climate policy stringency groups.

3.3. Meta-analytical procedures

Previous meta-analytical reviews on the CEP–CFP relationship were based on two different approaches, namely the aggregation technique of Hunter *et al* (1982) (hereafter: HS) (e.g. Orlitzky *et al* 2003, Albertini 2013) and the Hedges-Olkin-type meta-analysis (hereafter: HOMA) (e.g. Endrikat *et al* 2014, Busch and Lewandoski., 2018). Johnson *et al* (1995) compare meta-analytical techniques and observe HS does not very effectively correct biases in the effect sizes before deriving mean effect sizes. As we deem this of great importance for accuracy, we use HOMA and correct for individual study artefacts (e.g. overestimation of the population effect size in small sample studies). As a robustness check, we also employ the HS method.

To test the effect size distribution on homogeneity, we calculate the Q-statistic. This is a nonparametric test to assess the significance of the differences of two matched samples. Parametric tests are only reliable when the sample follows a normal distribution (Hunter *et al* 1982). A parametric test may yield significant results for the differences between the constructed subgroups. However, since the effect sizes in a small sample usually are not normally distributed, a non-parametric test is more informative. In this regard, a significant Q indicates a heterogeneous distribution and suggests the presence of moderating variables (Tavakol 2018). In line with Hedges and

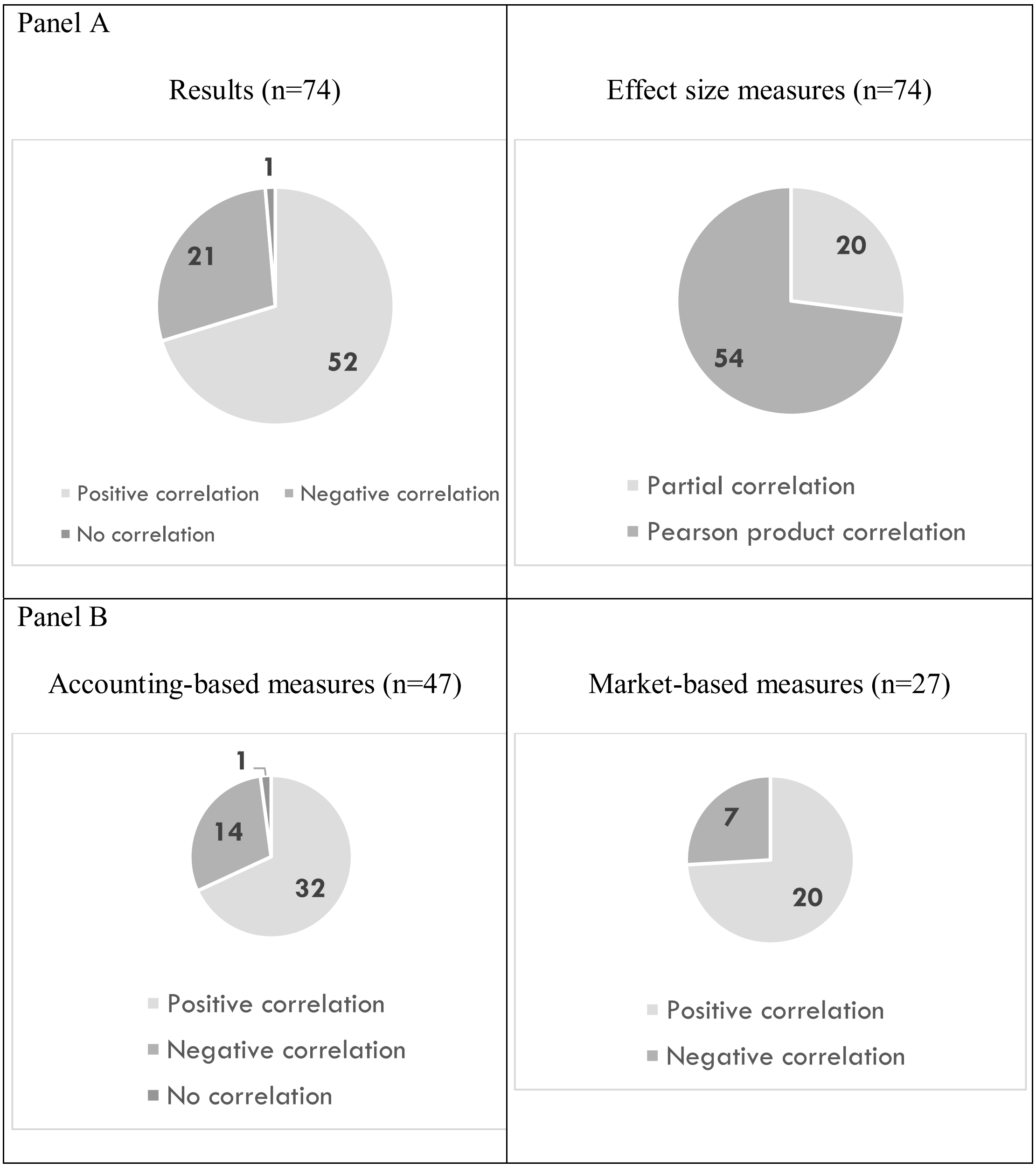

Figure 1. Key characteristics of the study sample.
Note: Figure 1 shows the key characteristics of our sample. Panel A gives information on the included effect sizes in this study. A total of 52 show positive effect. There are 21 negative effect sizes, and in one case no effect relationship was observed. A total of 54 effect sizes are gathered using Pearson product-correlations, and 20 are based on partial correlation coefficients. Panel B details the two CFP measures in the 34 sample studies. A total of 47 effect sizes are measured using accounting-based indicators for CFP. From which 32 indicate a positive, 14 a negative, and one measures no relationship. A total of 27 observations are measured using market-based indicators for CFP from which 20 measure a positive and 7 a negative relationship. Panel C shows how GHG performance is measured in the sample studies. A total of 47 effect sizes are measured using accounting-based indicators for CFP. From which 32 indicate a positive, 14 a negative, and one measures no relationship. A total of 27 observations are measured using market-based indicators for CFP from which 20 measure a positive and 7 a negative relationship. For Panels A-C, the sample consists of 34 studies, 74 effect sizes, and a total of 107 605 observations. Panel D reports the number of individual firms in different industries in the sample studies based on standard industrial classification. A total of 9716 firms are active in the manufacturing industry; 2210 individual firms are active in the service industry; 1875 firms are active in the transportation, communication, electric and gas and sanitary services industries; 1237 firms are active in the finance, insurance, and real estate industries; 1168 firms are active in the retail trade industry; 303 firms are active in the wholesale trade industry; 168 firms are active in the mining industry; 115 firms are active in the construction sector; 369 firms are active in the agriculture, forestry and fishing industries; from a total of 7074 firms, the industry is unknown. A total of 24 235 individual firm observations are included in the 34 sample studies.

Olkin (1985), we perform the Chi-square goodness of fit test with an alpha of 5% to test for the homogeneity of the distribution of the 74 effect sizes from the studies in table 2. The highly significant p-value ($pQ = 0.000$; see first line in table 3) indicates that the subgroups have different distributions and, therefore, there are likely to be moderating effects (Çoğaltay and Karadağ 2015). We are careful with interpreting the findings from subgroup analyses by using interaction tests as the analyses are not based on randomized

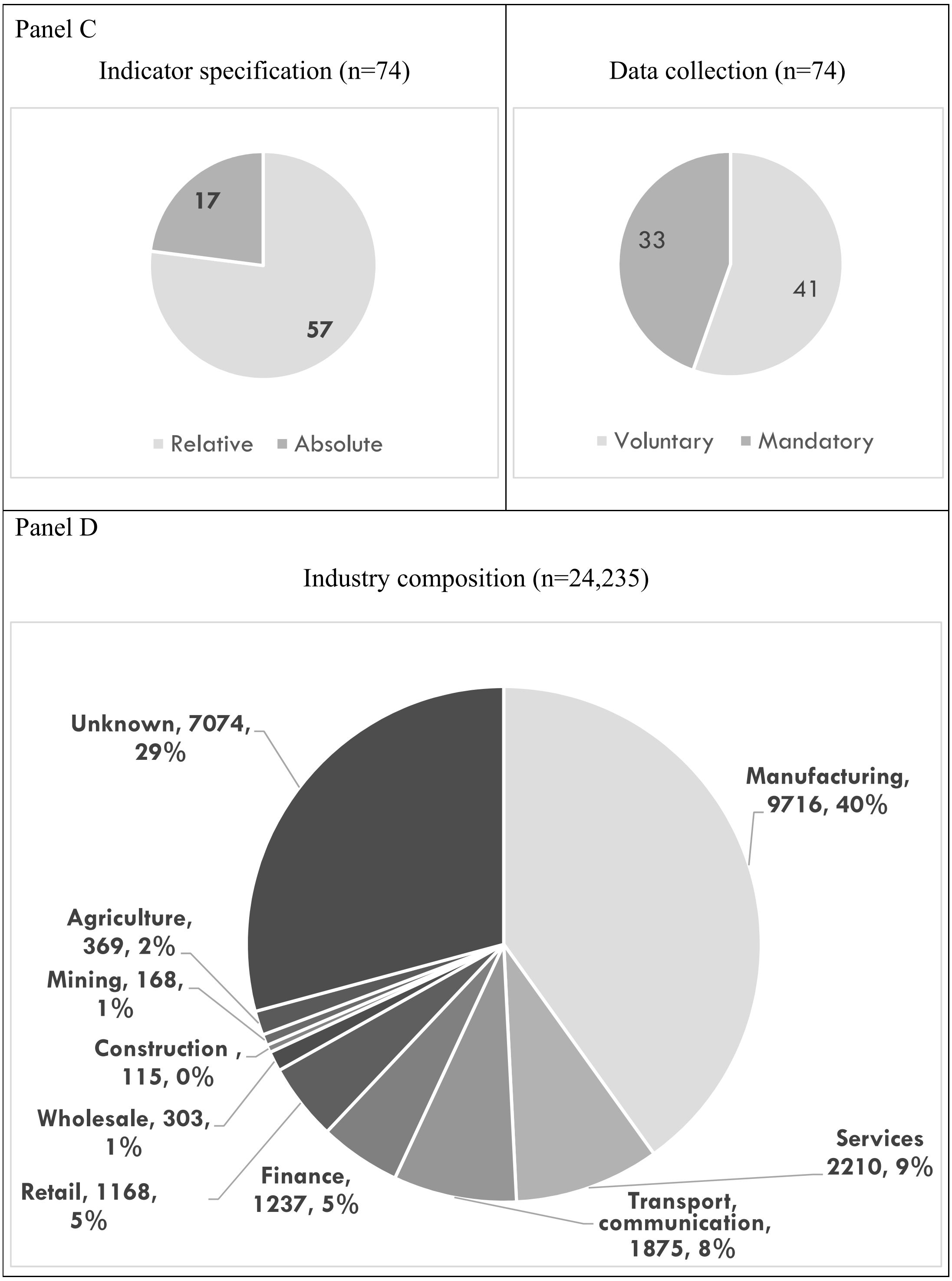

Figure 1. (Continued.)

groups of firms and therefore prone to confounding (Sedgwick 2015).

In addition, we want to test for the publication bias as studies with significant results have a higher probability of being published than studies with insignificant results. Here, we rely on the failsafe-N test of Rosenthal (1979). This test calculates the number of insignificant studies that should have to be included in the sample in order to arrive at an insignificant aggregated effect size (see Stanley and Doucouliagos 2012).

In order to test our hypotheses, several subgroups are constructed. We compare the subgroups to study whether the defining issue for classification indeed is relevant in relation to heterogeneity in our sample (see also Hedges and Olkin 1985). To determine whether the heterogeneity between subgroups is statistically significant, we also calculate Cochran's Q score and corresponding p-value using the Chi-square goodness of fit test. Because the sample size of the study is relatively small, it is important to realize that the Q statistic may provide

Table 3. Results meta-analysis.

	k	N	r	95% CI	Z	p	Q	pQ	pQ-bet	pMWW
Overall	74	107 605	0.052	0.02–0.08	3.47	0.001	984.60	0.000		
Moderators										
Reporting type									0.498	0.270
Mandatory	41	84 175	0.04	0.01–0.08	2.52	0.012	713.02			
Voluntary	33	23 430	0.07	0.02–0.12	2.55	0.011	244.01			
CEP indicator specification									0.207	0.550
Absolute	17	17 247	0.09	0.02–0.16	2.52	0.012	495.86			
Relative	57	90 358	0.04	0.01–0.07	2.29	0.022	449.46			
CFP measures									0.458	0.755
Market-based	28	49 045	0.07	0.02–0.12	2.43	0.015	649.36			
Accounting based	46	58 560	0.04	0.01–0.08	2.51	0.012	344.03			
Pollution intensity of industry									0.138	0.014
Only pollution-intensive industries	25	23 219	0.04	−0.01–0.08	1.56	0.119	136.05			
Mixed industries	25	23 580	0.08	0.04–0.13	3.68	0.000	112.85			
No data on industries	22	60 806	0.04	−0.03–0.10	1.14	0.252	716.85			
ETS stringency C3I CPI									0.558	
High	20	19 045	0.09	0.05–0.12	4.48	0.000	35.25			
Medium	15	17 919	0.05	−0.02–0.13	1.34	0.181	108.56			
Medium-low	10	7577	0.02	−0.08–0.12	0.41	0.680	102.36			
Low	9	7783	0.05	−0.04–0.14	1.19	0.234	60.03			
Global/no data on country	20	55 281	0.03	−0.02–0.08	1.21	0.223	619.12			
ETS stringency CAT									0.517	
Sufficient	24	20 138	0.09	0.03–0.15	2.84	0.005	145.52			
Moderate	15	16 132	0.05	−0.02–0.11	1.41	0.159	69.87			
Inadequate	16	16 805	0.04	−0.02–0.10	1.40	0.162	89.40			
Global/no data	21	56 032	0.03	−0.02–0.08	1.21	0.223	619.12			
Robustness check									0.156	0.068
Correlation	53	96 329	0.04	0.01–0.07	2.29	0.022	855.40			
Partial-correlation	21	11 276	0.09	0.03–0.15	2.82	0.005	166.28			

Note: Table 3 summarizes the results of the meta-analysis based on the Hedges and Olkin (1985) method. It first gives the overall aggregated relationship between corporate GHG performance and CFP. Next, it shows the results of the different subgroup analyses. It gives the aggregated effect sizes for the subgroups for different reporting types and the indicator specification of the corporate GHG performance construct. Further, it reports the effect sizes for the market and accounting based CFP indicator specification and the industry carbon intensity. It also reports the ETS stringency hypothesis using two different methods. For ETS stringency based on the C3-I and CCPI, the 'high' group consists seven studies conducted in the most stringent environments, the following seven studies from subsequently lower ETS stringency environments form the group 'medium,' the seven following studies form the group 'medium-low and the studies conducted in the lowest ETS stringent regions studies form the group low. The CAT ETS stringency measure has resulted in three groups, studies which for the group 'sufficient' are from regions with sufficient policies for reaching the UN climate goals. Moderate forms the group of studies which are performed in countries with moderate policies, and inadequate forms the groups of studies which are performed in inadequate performing countries. For the group 'global/no data available,' the study was conducted globally, or no information about the studied country was. k = number of effect sizes; N = total sample size; r = aggregated effect size 95% CI = 95% confidence intervals for the aggregated effect sizes; p = probability Q = Q statistic Pq = probability of Q statistic pQ-bet = p-value of between-group heterogeneity, p-MWW probability Mann–Whitney–Wilcoxon test (the MWW tests for ETS policy stringency are reported in appendix B.

a misleading measure of heterogeneity and should be interpreted with care (Sedgwick 2015, Tavakol 2018). To address this issue and to test whether subgroups differ significantly from one another since the effect-sizes of subgroups are unpaired, we also perform the non-parametric Mann–Whitney–Wilcoxon test. This test does not assume normally distributed or paired data (Fay and Proschan 2010). Here, the effect-sizes in the subgroups are not weighted, as differences in sample size would make the differences significant by definition.

4. Results

Table 3 presents the results from the meta-analysis for the relationship between corporate GHG performance and CFP. Regarding the overall effect, the aggregation of the effect sizes indicates a statistically significant positive relationship between GHG performance and CFP ($r = 0.05$, $Z = 3.47$, $p = 0.001$), based on a total of 74 effect sizes and 107 605 observations. This suggests that GHG performance is positively related to CFP. Therefore, we accept hypothesis 1A ('The overall relationship between corporate GHG performance and corporate financial performance is positive'). The significant positive association supports the eco-efficiency and stakeholder perspective and rejects the view of a trade-off between both constructs. It seems companies can improve their financial performance via the efficiency benefits of reducing their GHG emissions, which apparently satisfies the needs of their stakeholders (Hatakeda *et al* 2012, Trinks *et al* 2020). The Q score is highly significant and confirms the heterogeneity of the sample.

Table 3 also reports the results for the analysis of the various subgroups. It shows that when emissions are measured by voluntary reporting types, it is positively and significantly related to CFP ($r = 0.07$, $p = 0.01$); the same as when using mandatory reporting types ($r = 0.04$, $p = 0.01$) ($pQb = 0.498$). The Mann-Whitney-Wilcoxon test also indicates that the subgroups do not differ significantly from each other ($p = 0.270$). Therefore, we reject hypothesis 2 ('the type of reporting scheme used influences the results in the GHG and CFP literature').

Further, table 3 shows that GHG performance is significantly positive related to CFP ($r = 0.09$, $p = 0.01$) when absolute GHG emissions are used. At the same time, it shows that relative GHG indicators are significant too ($r = 0.04$, 0.01). Here, $pQb = 0.207$, and the Mann–Whitney–Wilcoxon analysis also shows that the differences between these two subgroups are not statistically significant ($p = 0.550$). As such, hypothesis 3B ('GHG performance affects CFP more when it is measured using absolute emissions compared to relative ones') is rejected.

Although the relationship between CEP and CFP is positive for both accounting- and market-based indicators, it appears to be somewhat stronger when market-based measures are used ($r = 0.07, p = 0.015$), than with accounting measures ($r = 0.04$, $p = 0.012$). However, we find an insignificant difference between these two groups ($pQb = 0.458$). In addition, the Mann-Whitney-Wilcoxon test results also suggest the difference is not statistically significant ($p = 0.755$). Hence, hypothesis 4A (GHG performance is more positively related to prior market-based than to prior accounting-based CFP) is rejected, as is its counterpart (4B).

Taking the industry perspective, table 3 show that studies that only included pollution-intense industries report lower effect sizes ($r = 0.04$, $p = 0.119$) than those with multiple industries ($r = 0.08$, $p = 0.00$). But the former is not significant and, hence, only in the mixed industry, GHG performance is significantly related to CFP. Based on the pQbet of 0.138 the two do not seem to differ in a statistically significant way. But the Mann–Whitney–Wilcoxon test results indicate that the differences between the subgroups are significant ($p = 0.014$). Based on the first test, we reject H5 (industry carbon intensity moderates the relationship between GHG and CFP; the GHG–CFP relationship is stronger in more polluting industries). However, on the basis of the Mann–Whitney–Wilcoxon test it appears that the GHG–CFP relationship seems to be significantly weaker for studies conducted in pollution-intense industries than for studies conducted in multiple industries. An explanation could be that over the years, forced by gradually tighter regulation, pollution-intense industries have already picked the 'low hanging fruits' (see also Delmas *et al* 2015).

For climate policy stringency, we first look into the way this is measured with the help of the CCPI and C3-I indices. In this regard, the relationship between GHG performance and CFP appears strongest for studies performed in countries with the most stringent policy regime ($r = 0.09$, $p = 0.00$). The CEP–CFP relationship for countries with medium-high stringency is insignificant ($r = 0.05$, $p = 0.18$), as is the case for sample countries in the medium-low cohort ($r = 0.02$, $p = 0.68$). For studies about countries that score lowest on policy stringency, the relationship also is insignificant ($r = 0.05$, $p = 0.23$). The Mann–Whitney–Wilcoxon test (reported in appendix B) demonstrates marginally significant differences between subgroups high and medium-high ($p = 0.089$), high and medium-low ($p = 0.095$), and significant differences between high and low ($p = 0.048$). This suggests that the GHG–CFP relationship is stronger in the most stringent climate policy regions. Next, we discuss the results based on CAT information. Here it shows that studies conducted in countries with policies qualified as sufficient show a clear positive and significant relationship between GHG and CFP ($r = 0.09$ $p = 0.005$). For the other subgroups, it is not significant. The results from the Mann–Whitney–Wilcoxon tests (see appendix B) reveal that most subgroups are not significantly different from one each other, with the exception of the group sufficient versus medium and insufficient combined. Therefore, hypothesis 6 ('the relationship between GHG performance and CFP is stronger for firms operating in countries with more stringent climate policy than for firms in countries with weak policy stringency') cannot be accepted on the basis of CAT information. The results suggest that the relationship between GHG and CFP is significant and positive for all subgroups, but is only significantly more so for the most climate policy stringent environments. This might be the case because initial phases of ETSs are characterized by low stringency, high bureaucracy, and little influence on innovation (Czerny and Letmathe 2017). These early phases are known for the free allocation of emission rights, low emission prices, and many industries being excluded (Abrell *et al* 2011).

In order to assess the reliability of the results of the meta-analysis, two robustness tests are performed: we use a different methodology and we rely on an alternative calculation of effect sizes. In addition, we account for the publication bias. First, we use the Hunter *et al* (1982) method to test the robustness of the HOMA analysis. This procedure is briefly explained in appendix C, and the results are in table C1 therein. It shows that the Hunter *et al* (1982) method yields qualitatively highly similar results to the HOMA method. The main difference is that it suggests there is a marginal significant difference between highly polluting industries and multiple industries, and between the different correlation coefficients. Second, in line with Hunter *et al* (1982), effect sizes were calculated for both correlations and estimated partial correlations (last row in table 3). The effect sizes measured based on correlation coefficients tend to be slightly higher ($r = 0.08$, $p = 0.00$) than effect sizes which were estimated based on partial-correlations ($r = 0.06$, $p = 0.004$). According to the Q-statistic, these correlations are not significantly different from each other ($pQ\text{bet} = 0.156$). However, the results from the MWW-test hint at marginally significant differences ($p = 0.068$). We also account for the presence of a publication bias (Rosenthal 1979). Here, the failsafe-N is calculated, which points at just very moderate existence of the publication bias. In particular, we find that 5576 (Z-score of 14.37) additional null-effect studies are required to make the summary effect size insignificant. This result can be explained by the fact that this study only includes studies that investigate the relationship between GHG emissions and CFP, and the number of studies on the topic is growing but still limited (Chapple *et al* 2011).

5. Conclusion

We conduct a review of the nascent literature after the relationship between companies' GHG emissions and financial performance. We employ a meta-analysis to examine whether there is a relationship between firms' GHG emissions and financial performance, what it looks like, and how sensitive the relationship is for research design and measurement. We investigate the results of studies undertaken after the signing of the Kyoto Protocol, as we regard this as a breakpoint in international climate policy. Hence, we focus on international studies for the period 1997–2019. We select peer-reviewed published academic studies using PRISMA sampling and end up with 34 relevant studies, including 74 effect sizes covering 107 605 observations. We observe that there are several drawbacks in the studies that relate physical and economic performance. In particular, it shows that the interaction mechanisms are not always described and motivated in a clear and coherent manner. Further, the measurement of both GHG emissions and financial performance in many cases is not transparent. In particular, it appears that not all studies clearly report how these emissions are being calculated and whether scope 1, scope 2, or scope 3 emissions are used. The required homogeneity of samples does not seem to be fully satisfied and there appears to be multiplicity. This potential of the multiplicity of data in the sample studies may lead to variability in the results. We observe that in many cases the sample studies do not clearly detail their procedure regarding the selection of countries, industries, and firms or the period studied. This is problematic and requires disciplining in this regards within the field of business and economics as it does not allow for full replication of the results.

Given these reflections and data limitations, the main finding of our study is that there is a significant positive relationship between companies' GHG performance and their financial performance, suggesting that companies with less GHG emissions show superior financial performance. Although the type of pollution is very different from other pollutants, this finding is in line with studies on the generic corporate environmental-financial performance relationship (e.g. Albertini 2013, Dixon-Fowler *et al* 2013, Endrikat *et al* 2014), as well as with a related study after the association between firms' carbon emissions and their financial performance (Busch and Lewandowski 2018). There are several ways to come to grips with both financial performance and GHG emissions. However, the choice of proxies for both does hardly appear to influence the results. For example, we establish that there is no significant difference when voluntary or mandatory GHG reporting information is used, when absolute or relative GHG emission measures are used, or when market or accounting based financial indicators are employed. However, this conclusion is based on a sample of studies that are hampered by problematic homogeneity and multiplicity. Therefore, we need to await further research to check for its reliability. Further, although there is some evidence that firms in less polluting industries outperform, we do not find substantial evidence that industry affiliation per se is a defining vector in the relationship between GHG emissions and financial performance. Looking into climate policy stringency, it appears that only in countries with the most stringent ETS regime, the relationship between emissions performance and financial performance is significantly more positive than elsewhere. We want to point out though that most sample studies focus on industrialized countries and suggest to study emerging markets and low income countries too. Our findings appear to be quite robust. This also is established by using an alternative meta-analytical procedure. Furthermore, we find there is no substantial publication bias. Therefore, on the basis of this review, we conclude there is a positive association between companies' GHG emission performance and their financial performance. In particular, companies with relatively low GHG emissions have relatively high financial performance.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

We highly appreciate the constructive comments of three anonymous reviewers. We thank faculty of the Stockholm Resilience Center, University of Groningen, and University of St Andrews for their feedback on previous versions of this manuscript. No external funding was provided for this study. The usual disclaimer applies.

Appendix A. Effect sizes

	Name of individual effect size	Effect sizes	Number of observations
1	Aggarwal and Dow (2012) ROA	0.31	325
2	Aggarwal and Dow (2012) TBQ	0.20	325
3	Brouwers *et al* (2018) ROA	−0.06	2593
4	Brouwers *et al* (2018) ROE	−0.03	2593
5	Brouwers *et al* (2018) TBQ	−0.05	2593
6	Brzobahat and Jansky (2010) Inca	−0.13	375
7	Brzobohatý and Jansky (2010) Inpa	−0.02	270
8	Brzobohatý and Jansky (2010) Inra	0.13	375
9	Busch *et al* (2012) CF/A	0.08	8089
10	Busch *et al* (2012) D/A	0.15	8089
11	Busch *et al* (2012) SR	0.01	8089
12	Busch *et al* (2012) TMR	0.03	8089
13	Busch *et al* (2012) USR	0.04	8089
14	Busch and Hoffmann (2011) ROA	−0.07	174
15	Busch and Hoffmann (2011) ROE	−0.08	174
16	Busch and Hoffmann (2011) TBQ	0.16	174
17	Chakrabarty and Wang (2013) ROE	0.05	264
18	Chakrabarty and Wang (2013) Sales	0.02	259
19	Chapple *et al* (2013) VE	0.30	58
20	Chapple *et al* (2013) VEMIT	0.28	58
21	Clarkson *et al* (2015) ROA	0.19	51
22	Clarkson *et al* (2015) TBQ	0.11	51
23	Clarkson *et al* (2015) ROA	0.11	842
24	Dangelico and Pontradolfo (2015) MP	0.32	122
25	Delmas *et al* (2015) ROA	0.39	3316
26	Delmas *et al* (2015) TBQ	0.00	2678
27	Fujii *et al* (2013) ROA	0.09	758
28	Fujii *et al* (2013) ROS	−0.07	758
29	Gallego-Alvarez *et al* (2014) ROA	0.05	3420
30	Gallego-Alvarez *et al* (2015) ROA	−0.02	267
31	Gallego-Alvarez *et al* (2015) ROE	0.11	267
32	Griffin *et al* (2017) PRCC	0.00	2235
33	Hatakeda *et al* (2012) prof	−0.09	1089
34	Hatakeda *et al* (2012) prof	−0.01	1089
35	Iwata and Okada (2011) ROA	0.09	751
36	Iwata and Okada (2011) ROE	0.03	751
37	Iwata and Okada (2011) ROI	0.13	751
38	Iwata and Okada (2011) ROIC	0.12	751
39	Iwata and Okada (2011) ROS	0.02	751
40	Iwata and Okada (2011) TBQ	0.07	749
41	Jung *et al* (2016) COD	0.02	225
42	Kim *et al* (2015) COE	0.10	1895
43	Kim *et al* (2015) ROA	0.03	1895
44	Kuo *et al* (2010) NIemissionreduction	0.33	32
45	Kuo *et al* (2010) NItotalemission	0.05	32
46	Lannelongue *et al* (2015) Profit	0.41	160
47	Lannelongue *et al* (2015) ROA	0.18	160
48	Lannelongue *et al* (2015) ROE	−0.01	160
49	Lee and Min (2015) ROA	0.03	2557
50	Lee *et al* (2015) TBQ	0.06	2557
51	Luo and Tang (2014) MR	0.02	336
52	Makridou *et al* (2019) CR	0.05	3952
53	Makridou *et al* (2019) EBITDA	0.03	3950
54	Makridou *et al* (2019) SR	0.07	3952
55	Matsumure *et al* (2014) MKT	−0.15	550
56	Misani and Pogutz (2015) ROA	−0.09	766

(Continued)

	Name of individual effect size	Effect sizes	Number of observations
57	Misani and Pogutz (2015) ROE	−0.07	766
58	Misani and Pogutz (2015) ROS	−0.10	766
59	Misani and Pogutz (2015) TBQ	−0.07	766
60	Nishitani and Kokubu (2012) TBQ	0.05	1888
61	Rokhamawati *et al* (2015) ROA	−0.28	90
62	Secinaro *et al* (2020) ROE1	0.02	125
63	Secinaro *et al* (2020) ROE2	0.07	125
64	Secinaro *et al* (2020) ROE3	0.03	125
65	Saka and Oshika (2014) MVE	0.27	1094
66	Tatsuo (2010) ROA1	0.08	350
67	Tatsuo (2010) ROA2	0.41	560
68	Tatsuo (2010) ROA3	0.05	380
69	Trumpp and Guenther (2017) ROA	0.10	1179
70	Trumpp and Guenther (2017) ROAS	0.01	1182
71	Trumpp and Guenther (2017) TSR	0.00	1182
72	Trumpp and Guenther (2017) TSRS	−0.02	1179
73	Wang *et al* (2014) TBQ	−0.10	69
74	Qi *et al* (2014) ROA	−0.16	98

Note: This appendix gives an overview of all included effect sizes in the meta-study. The table gives information regarding all individual effect sizes and corresponding sample sizes. A total of 75 effect sizes are extracted from 34 individual empirical studies with 107 605 individual observations. ROA = Return on Assets; ROE = return on equity; ROS = return on sales; ROI = return on investment; ROIC = return on invested capital; Inpa = Profit over assets; Lnra = revenues over assets; Inca = cost over assets; PRCC = stock price three months after fiscal year-end CF = cashflow; EBITDA = Earnings before interest, tax, deductions, and amortization; V = market value of common equity; CT = capital turnover CoE = cost of equity; MKTE = Market value total equity MVE = market value equity; TSR = total stock return.

Appendix B. Mann–Whitney–Wilcoxon test results for ETS stringency measures

CAT	Insufficient	Medium	Sufficient	Insufficient + medium	
Insufficient	—	0.868	0.133	—	
Medium	0.868	—	0.106	—	
Sufficient	0.133	0.118	—	0.040	
CCPI-C3I	**Highest**	**Medium**	**Medium-low**	**Low**	**High**
High	—	0.089	0.080	0.048	
Medium	0.089	—	0.782	0.815	
Medium-low	0.095	0.782	—	0.539	
Low	0.048	0.815	0.539	—	
M-ML-L					0.018

Note: This appendix presents the results of the non-parametric Mann–Whitney–Wilcoxon tests that are performed to determine whether subgroups differed significantly from each other. The first part shows whether subgroups developed based on the Climate Action Tracker scores differed significantly from each other. The first three rows and columns measure whether the subgroups (based on the ETS stringency of the countries in which they are performed) differ significantly from each other. The last column shows the results of the test whether studies form insufficient and medium scoring groups differ significantly from studies performance in sufficient scoring countries. The second part of the table describes the results of the tests whether subgroups developed based on the Climate Change Cooperation Index and the Climate Change Performance Index differ significantly from each other. Group high consists of the seven studies conducted in countries with the most stringent ETS policies. The group medium, consists of the seven studies from the following most stringent ETS countries. The subgroup medium-low consists of the seven studies from the following most stringent ETS countries. The group low consists of the studies from countries from which the ETS policies are the least stringent. The last column tests whether the combined group highest and medium differ significantly from the group medium-low and low.

Appendix C. Hunter *et al* meta-analytical method

In contrast to the method of HOMA method, the Hunter *et al* method does not put emphasis on isolating and correcting sources of error and bias (Stanley and Doucouliagos 2012). The method uses the untransformed effect-sizes estimates, and weights are based only on the sample size (Field 2003). Mean effect sizes are calculated as follows:

$$\bar{r} = \frac{\sum_{i=1}^{K}(n_i * r_i)}{\sum_{i=1}^{K} n_i} \tag{C1}$$

The variance across sample effect sizes consists of the variance of the effect sizes of the population and the sampling error. As such, the variance in population effect sizes is calculated using the sampling error. The following equation is used to calculate the variance of the sample effect sizes

$$\sigma_{\mathrm{r}}^2 = \frac{\sum\limits_{i=1}^{K} n_i\left(r - \bar{r}_i\right)}{\sum\limits_{i=1}^{K} n_i}. \tag{C2}$$

The error variance of the sample is calculated as:

$$\sigma_{\mathrm{e}}^2 = \frac{\left(1 - \bar{r}^2\right)^2}{\bar{N} - 1}. \tag{C3}$$

The variance in population effect size is estimated by subtracting the sampling error variance from the effect size sample variance. It is calculated with the following equation:

$$\widehat{\sigma_{\mathrm{p}}^2} = \sigma_{\mathrm{r}}^2 - \sigma_{\mathrm{e}}^2. \tag{C4}$$

Credibility intervals, at the 95% level, are calculated by subtracting the square root of the population variance multiplied by 1.96:

$$\mathrm{CI}_{\mathrm{Upper}} = \bar{r} + 1.96 \times \sqrt{\sigma_{\mathrm{p}}^2} \tag{C5}$$

$$\mathrm{CI}_{\mathrm{Lower}} = \bar{r} - 1.96 \times \sqrt{\sigma_{\mathrm{p}}^2}. \tag{C6}$$

Table C1. Meta-analytical results based on the Hunter *et al* method.

	k	*N*	*r*	95% CI	*Z*	*P*	*Q*	*pQ*	*p*Qbet
Overall	74	107 605	0.05	0.03–0.07	4.05	0.000	984.60	0.000	
Moderators									
Reporting type									0.544
Mandatory	41	84 175	0.04	0.01–0.08	2.88	0.004	713.02		
Voluntary	33	23 430	0.06	0.02–0.10	2.92	0.004	244.01		
CEP indicator specification									0.237
Absolute	17	17 247	0.09	0.06–0.18	2.09	0.036	449.46		
Relative	57	90 358	0.04	0.02–0.06	3.69	0.000	495.86		
CFP measures									0.354
Market-based	28	49 045	0.07	0.02–0.11	2.80	0.005	649.36		
Accounting based	46	58 560	0.04	0.02–0.07	3.08	0.002	334.03		
Carbon intensive industry									0.060
Only Carbon Intensive industries	25	23 219	0.03	0.01–0.06	2.00	0.046	136.05		
Mixed industries	25	23 580	0.08	0.04–0.11	4.35	0.000	112.85		
No report on industries	22	60 806	0.04	−0.01–0.09	1.61	0.011	716.40		
ETS stringency C3I CPI									0.349
Group 1	20	19 045	0.08	0.05–0.10	6.08	0.000	35.25		
Group 2	15	17 919	0.04	−0.01–0.08	1.75	0.081	108.56		
Group 3	10	7577	0.02	−0.06–0.10	0.59	0.557	102.36		
Group 4	9	7783	0.05	−0.01–0.11	1.55	0.120	60.03		
Global/no data	20	55 281	0.03	−0.2–0.08	1.27	0.204	619.12		

(Continued)

Table C1. (Continued.)

	k	N	r	95% CI	Z	P	Q	pQ	pQbet
ETS stringency CAT									0.302
Sufficient	16	16 805	0.08	0.04–0.12	3.81	0.000	145.52		
Moderate	15	16 132	0.04	0.00–0.08	1.89	0.059	69.87		
Inadequate	24	20 138	0.04	0.00–0.08	1.72	0.085	91.25		
Global/no data	21	56 032	0.03	0.01–0.08	1.27	0.204	619.12		
Robustness check									0.076
Correlation based	53	96 329	0.04	0.01–0.07	2.57	0.010	855.40		
Partial correlation based	21	11 276	0.08	0.00–0.14	3.62	0.000	116.28		

Note: This table summarizes the meta-analytical results based on the Hunter *et al* (1982) method to test the robustness of the Hedges and Olkin (1985) results. It first describes the overall aggregated relationship between corporate GHG performance and CFP. Next, the results of the different subgroups analyses are presented: The reporting type and the indicator specification of the corporate GHG performance construct are given, the CFP indicator specification and the industry carbon intensity are reported, the ETS stringency hypothesis using two different methods is reported. For ETS stringency based on the C3-I and CCPI, the seven most stringent studies form the 'high' group, followed by the next seven studies which are the group 'medium', the seven following studies form the group 'medium-low and the studies conducted in the lowest ETS stringent regions studies form the group low. The CAT ETS stringency measure has resulted in 3 groups, studies which for the group 'sufficient' are performed in a country with sufficient policies for reaching the UN climate goals. Moderate forms the group of studies which are performed in countries with moderate policies for reaching the UN climate goals and inadequate forms the groups of studies which are performed in poor performing countries. For the group 'global/no data available,' the study was conducted globally or no information about the included countries was available. Differences based on the use correlation coefficient or partial correlation are described to test the robustness of the results. k = number of effect sizes; N = total sample size; r = aggregated effect size 95% CI = 95% confidence intervals for the aggregated effect sizes; p = probability Q = Q statistic Pq = probability of Q statistic pQbet = p value of between group heterogeneity

Appendix D: ETS Stringency

Table D1. Key characteristics of the three stringency measures.

	Climate change performance index	Climate change cooperation index	Climate action tracker
Definition	The CCPI tracks countries' efforts to combat climate change. The score is based on total emissions, renewable energy use, and climate policies (Burck *et al* 2016).	The C3-I captures the overall performance as well as the performance in terms of political behavior and emissions; it allows for a global comparison of the climate policies of countries (Bernauer and Böhmelt 2013).	The CAT tracks the historical emissions and climate actions of countries towards the globally agreed aim of holding global warming below 2 °C, and pursuing efforts to limit warming to 1.5 °C (New Climate; Climate Analytics 2011).
Number of countries included	58	172	61
Time-period covered	2005–2019	1996–2008	2011–2019
Policy component assessed by	Expert assessment	Observed behavior	Proposed future emission cuts
Weight of emissions in the index	80%	50%	100%

Table D2. Studies ranked in relation to environmental policies.

Study number	Subgroups based on	C3I & CPI		CAT
		Rank	Group	Group
4	Busch and Hoffmann (2011)	—	—	—
31	Trumpp and Guenther (2017)	—	—	—
5	Busch *et al* (2012)	—	—	—
12	Delmas *et al* (2015)	—	—	—
27	Misani and Pogutz (2015)	—	—	—
15	Gallego-Álvarez *et al* (2015)			
26	Matsumura *et al* (2014)	29	M	Sufficient
16	Griffin *et al* (2017)	28	L	Insufficient
1	Aggarwal and Dow (2012)	27	L	Insufficient
6	Chakrabarty and Wang (2013)	26	L	Insufficient
14	Gallego-Alvarez *et al* (2014)	25		
32	Wang *et al* (2014)	24	H	Insufficient
24	Luo and Tang (2014)	23	M	Insufficient
20	Jung *et al* (2016)	22	L	Insufficient
29	Rokmawati *et al* (2015)	21	M–L	Sufficient
30	Saka and Oshika (2014)	20	M–L	Sufficient
34	Qi *et al* (2014)	19	L	Insufficient
21	Kim *et al* (2015)	18	M–L	Sufficient
17	Hatakeda *et al* (2012)	17	M–L	Moderate
9	Clarkson *et al* (2015)	16	H	Insufficient
23	Lee *et al* (2015)	15	M	Sufficient
22	Lannelongue *et al* (2015)	14	M	Moderate
11	Dangelico and Pontrandolfo (2015)	13	H	Sufficient
28	Nishitani and Kokubu (2012)	12	M	Sufficient
13	Fujii *et al* (2013)	11	M-L	Sufficient
2	Brouwers *et al* (2018)	9	M	Insufficient
3	Brzobohatý and Jansky (2010)	8	M-L	Medium
7	Chapple *et al* (2013)	7	H	Sufficient
19	Iwata and Okada (2011)	6	L	Sufficient
18	Kuo *et al* (2010)	5	H	Sufficient
34	Tatsuo (2010)	4	H	Sufficient
24	Makridou *et al* (2019)	3	H	Medium
30	Secinaro *et al* (2020)	2	H	Medium
8	Clarkson *et al* (2015)	1	H	Insufficient

Note: Table D2 shows the ranks of sample studies based on the ETS stringency of the countries in which they are performed. The table shows the study numbers and the ranking based on the Climate Change Cooperation Index and the Climate Change performance index. Group 'high' contains the seven highest-ranked studies; group 'medium' the next highest scoring seven studies; group 'medium-low' the following ranked seven studies; group 'low' contains the seven lowest-ranked studies. For the first six studies in the table, no ranks can be calculated. The table also shows the ranking based on the Climate Action tracker. The policy of the country of sample studies can be ranked sufficient, medium, or insufficient. A total of 28 studies was ranked.

Appendix E: Methodological issues

Before calculating the summary effects, the effect sizes are transformed to a standard normal metric by Fisher's z transformation to address skewness (see Hedges and Olkin 1985) using the following formula:

$$Zi = \frac{1}{2} \times \mathrm{Log}_e \frac{1+ri}{1-ri} \tag{E1}$$

where Z is the transformed partial correlation and r is the correlation coefficient. In line with Hedges and Olkin (1985), the weight assigned to the individual effect sizes is a variance component that consists of both the between-study and the within-study variance. The within-study variance V_{within} is:

$$V_{\text{within}} = \frac{1}{ni-3}. \tag{E2}$$

The between-study variance V_{between} is:

$$V_{\text{between}} = \frac{Q-(k-1)}{C} \tag{E3}$$

where

$$Q = \sum_{i=1}^{K} w_i Z_i - \frac{\left(\sum_{i=1}^{K} w_i Z_i\right)^2}{\sum_{i=1}^{K} w_i} \tag{E4}$$

$$w_i = \frac{1}{V_{\text{within}}} \tag{E5}$$

$$C = \sum_{i=1}^{K} W_i - \frac{\sum_{i=1}^{K} w_i}{\sum_{i=1}^{K} w_i}. \tag{E6}$$

The random-effect aggregated effect size is calculated by using the sum of the between-study and the within-study variance, V (Hedges and Olkin 1985):

$$V_{\text{i}} = V_{\text{within}} + V_{\text{between}}. \tag{E7}$$

In line with Hedges and Olkin (1985), we assign weights to each effect size based on the inverse value of the sum of the between and within-study variance by the following equation:

$$W_i = \frac{1}{V_i}. \tag{E8}$$

The mean effect size and the standard error of the mean effect size are calculated in line with Hedges and Olkin (1985) using the following equations

$$\bar{z}_{\text{r}} = \frac{\sum_{i=1}^{K} (W_i * Z_i)}{\sum_{i=1}^{K} W_i} \tag{E9}$$

$$\text{SE}(\bar{z}_{\text{r}}) = \sqrt{\frac{1}{\sum_{i=1}^{K} W_i}}. \tag{E10}$$

The confidence interval for the aggregated effect size is calculated by

$$\text{CI}_{\text{Upper}} = \bar{z} + 1.96 \times \text{SE}(\bar{z}_{\text{r}}) \tag{E11}$$

$$\text{CI}_{\text{Lower}} = \bar{z} - 1.96 \times \text{SE}(\bar{z}_{\text{r}}). \tag{E12}$$

Further, all values are transformed back to correlation units using

$$r_i = \frac{e^{2zi}-1}{e^{2zi}+1} \tag{E13}$$

Publication bias

To test for the publication bias, we calculate the failsafe-N (see Rosenthal 1979). The failsafe-N test calculates the number of insignificant studies that have to be included in the sample to make the aggregated effect size statistically insignificant (see Stanley and Doucouliagos 2012). The number of additional scores that have to be included to make the aggregated effect size insignificant at the 5% level is calculated as follows:

$$k \times \left[\frac{Z_{\text{s}}}{Z_{\text{a}}}\right]^2 - k \tag{E14}$$

where Z_{a} is the critical upper-tail value of the normal distribution, and Z_{s} is calculated as follows:

$$Z_{\text{s}} = \frac{\left(\sum_{i=1}^{K} Z_{\text{scores}}\right)}{\sqrt{k}} \tag{E15}$$

ORCID iD

Bert Scholtens https://orcid.org/0000-0001-5774-5191

References

Abrell J, Anta Ndoye F and Zachmann G 2011 Assessing the impact of the EU ETS using firm level Bruegel working paper 2011/08

Aggarwal R and Dow S 2012 Corporate governance and business strategies for climate change and environmental mitigation *Eur. J. Finance* **18** 311–31

Albertini E 2013 Does environmental management improve financial performance? A meta-analytical review *Organ. Environ.* **26** 431–57

Alkhurst M, Morgheim J and Lewis R 2003 Greenhouse gas emissions trading in BP *Energy Policy* **31** 657–63

Allouche José and Laroche Patrice 2005 A Meta-analytical investigation of the relationship between corporate social and financial performance *Revue de Gestion des Ressources Humaines* (Eska) p 18

Ambec S and Lanoie P 2008 Does it pay to be green? A systematic overview *Acad. Manage. Perspect.* **22** 45–62
Arimura T H, Hibiki A and Katayama H 2008 Is a voluntary approach an effective environmental policy instrument? A case for environmental management systems *J. Environ. Econ. Manage.* **55** 281–95
Bansal P and Roth K 2000 Why companies go green: a model of ecological responsiveness *Acad. Manage. J.* **43** 717–36
Bernauer T and Böhmelt T 2013 National climate policies in international comparison: the climate change cooperation index *Environ. Sci. Policy* **25** 196–206
Böhringer C 2003 The Kyoto protocol: a review and perspectives *Oxford Rev. Econ. Policy* **19** 451–66
Botta E and Kozluk T 2014 Measuring environmental policy stringency in OECD countries OECD Economics Department Working Papers p 1177
Bowen A 2018 *What is a Carbon Price and Why Do We Need One?* (London: Grantham Research Institute)
Brouwers R, Schoubben F and van Hulle C 2018 The influence of carbon cost pass through on the link between carbon emission and corporate financial performance in the context of the European Union Emission Trading Scheme *Bus. Strategy Environ.* **27** 1422–36
Brzobohatý T and Jansky P 2010 Impact of CO_2 emissions reductions on firms' finance in an emerging economy: the case of the Czech Republic *Transit. Stud. Rev.* **17** 725–36
Burck J, Hagen U and Hoehne N 2016 *Climate Change Performance Index* (Bonn: Germanwatch)
Busch T and Hoffmann V 2011 How hot is your bottom line? Linking carbon and financial performance *Bus. Soc.* **50** 233–65
Busch T, Lehman N and Hoffmann V 2012 *Corporate Social Responsibility Negative Externalities and Financial Risk: The Case of Climate Change* (Amsterdam: Duisenberg School of Finance)
Busch T and Lewandowski S 2018 Corporate carbon and financial performance: a meta-analysis *J. Ind. Ecol.* **22** 745–59
Chakrabarty S and Wang L 2013 Climate change mitigation and internationalization: the competitiveness of multinational corporations *Thunderbird Int. Bus. Rev.* **55** 673–88
Chapple L, Clarkson M and Gold D 2013 The cost of carbon: capital market effects of the proposed emission trading scheme (ETS) *Abacus* **49** 1–33
Chapple L, Clarkson P and Overell M B 2011 Environmental reporting and its relation to corporate environmental performance *Abacus* **47** 27–60
Chen L and Gao L 2012 The pricing of climate risk *J. Financ. Econ. Pract.* **12** 115–31
Clarkson P, Li Y, Pinnuck M and Richardson G D 2015 The valuation relevance of greenhouse gas *Eur. Account. Rev.* **24** 551–80
Çoğaltay N and Karadağ E 2015 Introduction to meta-analysis *Leadership and Organizational Outcomes: Meta-analysis of Empirical Studies* Karadağ Engin (Springer) pp 19–28
Czerny A and Letmathe P 2017 Eco-efficiency: GHG reduction related environmental and economic performance. The case of the companies participating in the EU Emissions Trading Scheme *Bus. Strategy Environ.* **26** 791–806
Dam L and Scholtens B 2015 Towards a theory of responsible investing: on the economic foundations of corporate social responsibility *Resour. Energy Econ.* **41** 103–21
Dangelico R M and Pontrandolfo P 2015 Being 'green and competitive': the impact of environmental actions and collaborations on firm performance *Bus. Strategy Environ.* **26** 413–30
Danso A, Adamoko S, Amankwah-Amoa J, Owusu-Agyei S and Konadu R 2019 Environmental sustainability orientation, competitive strategy and financial performance *Bus. Strategy Environ.* **28** 885–95
Delmas M A, Nairn-Birch N and Lim J 2015 Dynamics of environmental and financial performance: the case of greenhouse gas emissions *Organ. Environ.* **28** 374–93
Delmas M, Hoffmann V H and Kuss M 2011 Under the tip of the iceberg: absorptive capacity, environmental strategy, and competitive advantage *Bus. Soc.* **50** 116–54
Deschenes O 2018 Environmental regulations and labor markets: balancing the benefits of environmental regulations for everyone and the costs to workers and firms *IZA World of Labor*
Dixon-Fowler H R, Slater D J, Johnson J L, Ellstrand A E and Romi A M 2013 Beyond 'does it pay to be green?' A meta-analysis *J. Bus. Ethics* **112** 353–66
Dowell G, Hart S and Yeung B 2000 Do corporate global environmental standards create or destroy market value? *Manage. Sci.* **46** 1059–74
Ekwurzel B, Boneham J, Dalton M W, Heede R, Mera R J, Allen M R and Frumhoff P C 2017 The rise in global atmospheric CO2, surface temperature, and sea level from emissions traced to major carbon producers *Clim. Change* **144** 579–90
Endrikat J, Guenther E and Hoppe H 2014 Making sense of conflicting empirical findings: a meta-analytic review of the relationship between corporate environmental and financial performance *Eur. Manage. J.* **32** 735–51
Enkvist P A, Nauclér T and Rosander J 2007 A Cost Curve for Greenhouse Gas Reduction (New York: McKinsey & Company)
Fay M and Proschan M 2010 Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules *Stat. Surv.* **4** 1–39
Field A 2003 Can Meta-analysis Be Trusted? *Psychologist* **16** 642–5
Friede G, Busch T and Bassen A 2015 ESG and financial performance: aggregated evidence from more than (2000) empirical studies *J. Sustain. Financ. Invest.* **5** 210–33
Fujii H, Iwata K, Kaneko S and Managi S 2013 Corporate environmental and economic performance of Japanese manufacturing firms: empirical study for sustainable development *Bus. Strategy Environ.* **22** 187–201
Gallego-Álvarez I, García-Sánchez I and da Silva Vieira C 2014 Climate change and financial performance in times of crisis *Bus. Strategy Environ.* **23** 361–74
Gallego-Alvarez I, Segura L and Martinez-Ferrero J 2015 Carbon emission reduction: the impact on the financial andoperational performance of international companies *J. Clean. Prod.* **103** 149–59
Griffin P A, Lont D H and Sun E Y 2017 The relevance to investors of greenhouse gas emission disclosures *Contemp. Account. Res.* **34** 1265–97
Guenther E, Hoppe H and Endrikat J 2011 Corporate financial performance and corporate environmental performance: a perfect match? *Z. Umweltpolitik Umweltrecht* **34** 154–84
Hart S L and Ahuja G 1996 Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance *Bus. Strategy Environ.* **30** 30–37
Hassel L, Nilsson H and Nyquist S 2005 The value relevance of environmental performance *Eur. Account. Rev.* **14** 41–61
Hatakeda T, Kokubu K Kajiwara T and Nishitani K 2012 Factors influencing corporate environmental protection activities for greenhouse gas emision reductions: the relationhip between environmental and financial performance *Environ. Resour. Econ.* **53** 445–81
Hedges L and Olkin I 1985 *Statistical Methods for Meta-Analysis* (New York: Academic)
Hunter J E, Schmidt F L and Jackson G B 1982 *Meta-analysis: Cumulating Research Findings across Studies* (Beverly Hills, CA : Sage)
IPCC 2014 *Climate Change 2014: Mitigation of Climate Change* (Cambridge: Cambridge University Press)
Iwata H and Okada K 2011 How does environmental performance affect financial performance *Ecol. Econ.* **70** 1691–700
Jiang R J and Bansal P 2003 Seeing the need for ISO 14001 *J. Manage. Stud.* **40** 1047e1067
Johnson B, Mullen B and Eduardo S 1995 Comparison of three major meta-analytic approaches *J. Appl. Psychol.* **80** 94–106

Joltreau E and Sommerfeld K 2018 Why does emissions trading under the EU Emissions Trading System (ETS) not affect firms' competitiveness? Empirical findings from the literature *Clim. Policy* **4** 453–71

Jung J, Herbohn K and Clarkson P 2016 Carbon risk, carbon risk awareness and the cost of debt financing *J. Bus. Ethics* **150** 1151–71

Kim Y, An H T and Kim J D 2015 The effect of carbon risk on the cost of equity capital *J. Clean. Prod.* **93** 279–87

King A A and Lenox M J 2001 Does it really pay to be green? An empirical study of firm environmental and financial performance *J. Ind. Ecol.* **5** 105–16

Kirca A and Yaprak A 2010 The use of meta-analysis in international business research: its current status and suggestions for better practice *Int. Bus. Rev.* **19** 306–14

Konar S and Cohen M 2001 Does the market value environmental performance? *Rev. Econ. Stat.* **83** 281–9

Kuik O and Mulder M 2004 Emissions trading and competitiveness: pros and cons of relative and absolute schemes *Energy Policy* **32** 737–45

Kuo L, Huang S K and Wu Y J 2010 Operational efficiency integrating the evaluation of environmental investment: the case of Japan *Manage. Decis.* **48** 1596–616

Lannelongue G, Gonzalez-Benito J and Gonzalez O 2015 Input, output, and environmental management productivity: effects on firm performance *Bus. Strategy Environ.* **24** 145–58

Lee K, Min B and Yook K 2015 The impacts of carbon (CO2) emissions and environmental research and development (R&D) investment on firm performance *Int. J. Prod. Econ.* **167** 1–11

Lewandowski S 2015 Carbon emissions and corporate financial performance: a systematic literature review and options for methodological enhancements *Corporate Carbon and Climate Accounting* eds S Schaltegger, D Zvezdov, I Alvarez Etxeberria, M Csutora and E Günther (Berlin: Springer) pp 193–215

Luo L and Tang Q 2014 Carbon tax, corporate carbon profile and financial return *Pac. Account. Rev.* **26** 351–73

Lyon T P and Maxwell J W 2011 Greenwash: corporate environmental disclosure under threat of audit *J. Econ. Manage. Strategy* **20** 3–41

Makridou G, Doumpos M and Galariotis E 2019 The financial performance affirms participating in the EU emissions trading scheme *Energy Policy* **129** 250–9

Matsumura E M, Prakash R and Vera-Munoz S 2014 Firm-value effects of carbon emissions and carbon disclosures *Account. Rev.* **89** 695–724

Misani N and Pogutz S 2015 Unraveling the effects of environmental outcomes and processes on financial performance: a non-linear approach *Ecol. Econ.* **109** 150–60

Moher D, Liberati A, Tetzlaff J and Altman D 2010 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement *Int. J. Surg.* **8** 336–41

New Climate; Climate Analytics 2011 The climate action tracker (available at: https://climateactiontracker.org/) (Accessed 22 November 2019)

Nishitani K and Kokubu K 2012 Why does the reduction of greenhouse gas emissions enhance firm value? The case of Japanese manufacturing firms *Bus. Strategy Environ.* **21** 517–29

Olsthoorn X, Wehrmeyer W, Tyteca D and Wagner M 2001 Environmental indicators for business: a review of the literature and standardisation methods *J. Clean. Prod.* **9** 453–63

Orlitzky M, Schmidt F L and Rynes S L 2003 Corporate social and financial performance *Organ. Stud.* **24** 403–11

Orsato R J 2006 Competitive environmental strategies: when does it pay to be green *Calif. Manage. Rev.* **48** 127–43

Perrault E and Clark C 2010 Should corporate social reporting be voluntary or mandatory? Evidence from the banking sector in France and the US *Corp. Govern.* **10** 512–26

Qi G, Zeng S, Shi J J, Meng X, Lin H and Yang Q X 2014 Revisiting the relationship between environmental and financial performance in Chinese industry *J. Environ. Manage.* **145** 349–56

Rokhmawati A, Sathye M and Sathye S 2015 The effect of GHG emission, environmental performance, and social performance on financial performance of listed manufacturing firms in Indonesia *Procedia Soc. Behav. Sci.* **211** 461–70

Rosenthal R 1979 The file drawer problem and tolerance for null results *Psychol. Bull.* **86** 638–41

Russo M V and Fouts P A 1997 A resource-based perspective on corporate environmental performance and profitability *Acad. Manage. J.* **40** 534–59

Saka C and Oshika T 2014 Disclosure effects, carbon emissions and corporate value *Sustain. Account. Manage. Policy J.* **5** 22–45

Sandoff A and Schaad G 2009 Does EU ETS lead to emission reductions through trade? The case of the Swedish emissions trading sector participants *Energy Policy* **37** 3697–977

Secinaro S, Brescia V, Calandra D and Saiti B 2020 Impact of climate change mitigation policies on corporate financial performance: evidence-based on European publicly listed firms *Corp. Soc. Responsib. Environ. Manage.* (https://doi.org/10.1002/csr.1971)

Sedgwick P 2015 Meta-analyses: what is heterogeneity? *BMJ* **350** h1435

Slawinski N, Pinkse J, Busch T and Banerjee S 2017 The role of short-termism and uncertainty avoidance in organizational inaction on climate change: a multi-level framework *Bus. Soc.* **56** 253–82

Stanley T and Doucouliagos H 2012 *Meta-regression Analysis in Economics and Business* (London: Routledge)

Sullivan R and Gouldson A 2012 Does voluntary carbon reporting meet investors needs? *J. Clean. Prod.* **36** 60–67

Tang S and Demeritt D 2018 Climate change and mandatory carbon reporting: impacts on business process and performance *Bus. Strategy Environ.* **27** 437–55

Tatsuo K 2010 An analysis of the eco-efficiency and economic performance of Japanese companies. Asian Business and Management 209–22

Tavakol M 2018 Making sense of meta-analysis in medical education research *Int. J. Med. Educ.* **10** 29–33

Tendal B, Nüesch E, Higgins J P T, Jüni P and Gøtzsche P C 2011 Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study *BMJ* **343** d4829

Trinks P J, Scholtens B and Mulder M 2020 An efficiency perspective on carbon emissions and financial performance *Ecol. Econ.* **175** 106632

Trinks P J, Scholtens B, Mulder M and Dam L 2018 Fossil fuel divestment and portfolio performance *Ecol. Econ.* **146** 740–8

Trumpp C and Guenther T 2017 Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance *Bus. Strategy Environ.* **26** 49–68

Van der Goot F and Scholtens B 2015 Carbon prices and firms' financial performance: an industry perspective *Carbon Manage.* **5** 491–505

Vishwanathan P, van Oosterhout J, Heugens P P M A R, Duran P and van Essen M 2020 Strategic CSR: a concept-building meta-analysis *J. Manage. Stud.* **57** 314–50

Waddock S and Graves S 1997 The corporate social performance–financial performance link *Strateg. Manage. J.* **18** 303–19

Wang L, Li S and Goa S 2014 Do greenhouse gas emissions affect financial performance?–An empirical examination of Australian public firms *Bus. Strategy Environ.* **23** 505–19

World Bank 2019 *State and Trends of Carbon Pricing 2019* (Washington, DC: World Bank Group)