

The Robustness of the Corporate Social and Financial Performance Relation: A Second-Order Meta-Analysis

Timo Busch^{1*} and Gunnar Friede^{1,2}

¹University of Hamburg, School of Business, Economics and Social Science, Hamburg, Germany

²Deutsche Asset Management Investment, Frankfurt am Main, Germany

ABSTRACT

For many decades, there has been a debate about the relation between corporate social/environmental performance (CSP) and corporate financial performance (CFP). Our study presents a review of academic research on this topic by applying a second-order meta-analysis. The data sample combines 25 previous meta-analyses yielding a sample size of one million observations. Our results demonstrate a highly significant, positive, robust, and bilateral CSP-CFP relation. The relation is positive regardless of whether firms focus on ecological or social aspects, though corporate reputation turns out to be a key CSP determinant. We find a particularly strong CSP-CFP relation for operational CFP. Furthermore, we add a new perspective on potential biases resulting from the studies' publishing source: social issues-oriented journals and methodological weaker papers do not distort the positive CSP-CFP relation. Our conclusion is: Based on the extant literature, the business case for being a good firm is undeniable. © 2018 The Authors. *Corporate Social Responsibility and Environmental Management* published by ERP Environment and John Wiley & Sons Ltd.

Received 24 April 2017; revised 22 October 2017; accepted 3 November 2017

Keywords: corporate social performance (CSP); corporate financial performance (CFP); meta-analysis; business case

Introduction

WHAT IS THE RELATION BETWEEN CORPORATE SOCIAL/ENVIRONMENTAL PERFORMANCE (CSP) AND CORPORATE FINANCIAL PERFORMANCE (CFP)? For nearly 50 years, scholars have investigated this question. The search for the answer has become particularly prominent since the end of the 1990s. Our study sample comprises 1214 empirical primary studies on the CSP-CFP relation. While considerable research has been conducted, it is as yet less clear how different CSP dimensions and CFP categories interrelate and in which way various moderating factors determine the CSP-CFP relation (Aguinis and Glavas, 2012; McWilliams & Siegel, 2001; Pelozo, 2009). Moreover, the robustness of the business case over time has only occasionally been analyzed (Borgers, Derwall, Koedijk, & ter Horst, 2013; Derwall, Koedijk, & Ter Horst, 2011). The aim of this study is to provide a more detailed understanding of the CSP-CFP relation.

*Correspondence to: Timo Busch, University of Hamburg, School of Business, Economics and Social Science, Hamburg, Germany.
E-mail: timo.busch@wiso.uni-hamburg.de

In 2003, the widely cited Orlitzky, Schmidt, and Rynes (2003) meta-study responded to calls for a moratorium on CSP-CFP research (Margolis & Walsh, 2001; Rowley & Berman, 2000). Orlitzky *et al.* (2003) found a significant, positive, and bidirectional correlation of CSP with CFP. Despite these findings, the authors claim that this relation cannot be generalized across all CSP dimensions and CFP categories, detecting a large variability – even after correcting for sampling error and measurement error. Since then, the picture has become even more ambiguous and critical debates have resurfaced about the actual CSP-CFP relation (Aldag, 2013; Barnett, 2007; Borgers *et al.*, 2013; Devinney, 2009; Fleming & Jones, 2013; Orlitzky, 2011, 2013; Wood, 2010). On the one hand, many scholars claim that, to date, the literature on the CSP-CFP relation remains inconclusive. On the other hand, a whole range of new studies and further meta-analyses have been published supporting a positive relation. Thus, more than a decade after Orlitzky *et al.* (2003), it seems to be the right time to revive the paper's spirit and ask: when generalizing existing research, is there still evidence that it pays off to be a good firm?

To answer this question, we apply a second-order meta-analysis. In other scientific disciplines like medicine, psychology, or education (i.e. research areas typically also experiencing vast amounts of data and conflicting results), meta-analytic techniques are widespread and have been applied since the 1960s (Glass, 1976; Hedges & Olkin, 1985; Hunter, Schmidt, & Jackson, 1982). With the growing number of meta-analyses, the idea of cumulative meta-analysis also appeared (Kazrin, Durac, & Agteros, 1979). Several labels have been used for such analyses: overview of reviews, review of reviews, umbrella reviews, meta-reviews, cumulative meta-analyses, meta-meta-analysis, and second-order meta-analysis (Cooper & Koenka, 2012). However, until recently, no standard procedure with comparable rigour to first-order meta-analytical approaches had emerged. This paper makes use of the novel method for conducting second-order meta-analysis proposed by Schmidt and Oh (2013).

We contribute to the academic inquiry on the CSP-CFP relation in several regards. By summarizing the findings of existing meta-analyses and utilizing a data sample that is more than 25 times larger compared to the average of previous studies, we are able to generate fairly robust results. Not only do we demonstrate that the main findings of Orlitzky and colleagues still hold today, but our results deliver more unequivocal and comprehensive conclusions. The highly significant, positive, and bilateral CSP-CFP relation varies in magnitude depending on different CSP dimensions and CFP categories, but effect sizes are less variable than expected. Surprisingly, the relation is positive regardless of whether firms focus on ecological or social aspects, though corporate reputation turns out to be a key aspect. We highlight that a particularly strong CSP-CFP relation exists when operational CFP is considered. Furthermore, we conclude that shrinking CSP-CFP correlations in meta-analyses over time can be ascribed to more sophisticated research methods and are not observable at primary study level. Finally, we address potential biases resulting from the studies' publishing source: we show that social issues-oriented and lower-impact journals as well as methodological weaker papers do not distort the CSP-CFP relation. The practical relevance of our second-order meta-analysis is: the business case for being a good firm is undeniable.

Hypotheses

Theoretical arguments support a positive, negative, as well as neutral CSP-CFP relation. Arguments supporting a positive CSP-CFP relation are typically rooted in stakeholder theory (Donaldson & Preston, 1995; Freeman, 1984), the resource-based view (Barney, 1991; Wernerfelt, 1984), institutional theory (DiMaggio & Powell, 1983), or draw on the industrial organization literature (Porter, 1979). When arguing for a negative CSP-CFP relation, scholars usually refer to Levitt's (1958) or Friedman's (1962, 1970) logic of social responsibility of the firm to solely increase its profits. Moreover, arguments based on portfolio theory (Markowitz, 1952, 1959) point towards disadvantageous limitations for portfolio construction when picking stocks following a CSP filter. Finally, proponents of the efficient market hypothesis (Fama, 1970, 1991) argue for a neutral – or at least non-exploitable – relation of CSP and CFP, as all information is priced at any time by the market. This also includes potentially business-relevant CSP information.

Next to these differing theoretical perspectives, the empirical CSP-CFP debate to date faces several questions. Even though considerable primary research has been conducted, it still remains less clear how different CSP dimensions and CFP categories interrelate and whether the business case holds as well for different measures of CFP. Eventually, questions of robustness and moderating effects arise which need to be answered in order to derive

unequivocal conclusions regarding the CSP-CFP relation and its manifestations. This paper seeks to answer some of these questions.

The general CSP-CFP Relation

Direction of the CSP-CFP

In the academic literature, to date, the business case for CSP seems to be increasingly accepted and empirically supported (Carroll & Shabana, 2010; Eccles, Ioannou, & Serafeim, 2014; Flammer, 2015). In line with the outcomes of previous research that summarized large numbers of primary research (e.g. Endrikat, Guenther, & Hoppe, 2014; Margolis, Elfenbein, & Walsh, 2009; Orlitzky *et al.*, 2003) we, thus, propose:

H1a: *CSP and CFP are generally positively related across a wide variety of study designs.*

Causality of the CSP-CFP Relation

Next to the general CSP-CFP relation, the subsequent question pertains to the causality of this relation. Proponents of good management theory (Waddock & Graves, 1997) argue that prior good CSP leads subsequently to better CFP. The opposite explanation is the slack resources theory (Cyert & March, 1963; J. B. McGuire, Sundgren, & Schneeweis, 1988), where excess financial resources of a company are seen as a necessary condition in order to re-invest and, thus, counter threats or exploit opportunities (Daniel, Lohrke, Fornaciari, & Turner, 2004). Following this view, prior superior CFP leads subsequently to improved CSP. As both theories found theoretical and empirical support, a common argument is that there exists a virtuous circle between CSP and CFP – they seem to be related reciprocally and concurrently (Orlitzky *et al.*, 2003; Hart & Ahuja, 1996; Waddock & Graves, 1997). We follow this line of thought and thus expect:

H1b: *The causality of the CSP-CFP relation is bidirectional.*

The Relation of various CSP Dimensions and CFP

Corporate Environmental Performance

Since the first review studies on the CSP-CFP relation (Aldag & Bartol, 1978; Arlow & Gannon, 1982), research has considered corporate environmental performance (CEP) as part of CSP. Orlitzky and Benjamin (2001) were the first to split the CSP construct and analyze CEP separately. These authors assume investors and consumers are less aware of the environmental performance of firms. In contrast, social aspects are more likely to be publicly communicated and, thus, reflected by stock- and stakeholders. Accordingly, the authors report a weaker CEP-CFP correlation compared to the overall sample. Orlitzky *et al.* (2003) reiterate the weak relation between CEP and CFP. Over the course of the years, individual meta-studies dedicated specifically to CEP and/or the social dimension of CSP were published. The summary effects for the CEP-CFP relation were often positive and significant (Albertini, 2013; Dixon-Fowler, Slater, Johnson, Ellstrand, & Romi, 2013; Endrikat *et al.*, 2014). However, the absolute level usually demonstrated lower correlation levels compared to studies purely focusing on social aspects (Combs, Liu, Hall, & Ketchen, 2006; Crook, Todd, Combs, Woehr, & Ketchen, 2011; Unger, Rauch, Frese, & Rosenbusch, 2011). In line with these findings, we expect:

H2a: *The relation between CSP (excluding environmental aspects) and CFP is stronger than for CEP and CFP.*

CSP Reputation

Four CSP dimensions attracted the greatest interest among meta-researchers in the past: CSP reputation, CSP disclosure, CSP philanthropy, and CSP audits, processes, and policies. CSP reputation (Barnett, Jermier, & Lafferty, 2006) typically yields high correlations with CFP. The reason being that corporate reputation is of particular

relevance or interest to a variety of external parties. For example, reputation is likely to influence customers' behaviours and investors' decisions (Orlitzky *et al.*, 2003). This intangible character of corporate reputation, which is difficult to replicate, leads to sustained superior profit over time (P. W. Roberts & Dowling, 2002). Literature reviews and meta-analytical results point in a similar direction (Margolis *et al.*, 2009; Orlitzky *et al.*, 2003; Sabate & Puente, 2003). We, thus, hypothesize that CSP reputation plays a dominant role:

H2b: *Within the CSP dimensions, reputational aspects have the strongest relation to CFP.*

CSP Disclosure

A more controversial debate exists regarding the link between CSP disclosure and CFP. Wiseman (1982) finds considerable discrepancies between CSP disclosures and actual CSP. Ullmann (1985) concludes that voluntary disclosure should not be considered as a proxy for CSP performance. R. W. Roberts (1992) discovers that stakeholder power, philanthropic activities, and previous economic performance significantly explain concurrent CSP disclosure. Orlitzky *et al.* (2003) ascribe a minimal positive relation between CSP disclosure practices and CFP, whereas additionally almost all of the observed effect size variance is due to statistical artifacts. Research in environmental management also nourishes a weak relation among both measures. Schultze and Trommer (2012, p. 396) summarize, based on a review of empirical research, that 'as disclosure is mostly voluntary, many companies only provide beneficial rather than accurate, unbiased information on their current performance'. This could lead, in turn, to distorted information of market participants accompanied by increased stock volatility when investors realize that their investment decisions were based on wrong assumptions (Orlitzky, 2013). Thus, we expect a relatively weaker link of CSP disclosure and CFP:

H2c: *Within the CSP dimensions, CSP disclosure has the weakest relation to CFP.*

The Relation of CSP and various CFP Categories

Perceptual CFP

Perceptual evaluations of business performance by senior executives are considered to be valid substitutes for objective CFP measures. The product-moment correlation coefficient between subjective and objective measures ranges in empirical studies between 0.4 and 0.8 (Guthrie, 2001; Venkatraman & Ramanujam, 1987; Wall *et al.*, 2004). Despite managers' discretionary assessment of the firm performance, the judgment is often a fair reflection of financial reality and not per se overrated (Venkatraman & Ramanujam, 1987). Depending on the research context, the application of survey-based measures can even be the preferred choice of researchers. A quarter of empirical studies in top management journals therefore use some kind of subjective organizational performance measure (P. J. Richard, Devinney, Yip, & Johnson, 2009). Perceptual CFP measures are compiled by one-fifth of the meta-studies in our sample. We formulate our hypothesis in accordance with meta-analyses that report high correlations between CSP and perceptual CFP measures (Margolis *et al.*, 2009; Orlitzky *et al.*, 2003; Wu, 2006):

H3a: *CSP has the strongest relation to perceptual CFP than to other CFP categories.*

Operational CFP

Operational performance such as productivity allows a nuanced conceptualization of business performance as it focuses on those operational success factors that might lead to superior CFP (Venkatraman & Ramanujam, 1986). As an intermediate-outcome metric (Peloza, 2009), it reflects information that is closer to the value creation process within a firm and more remote to the variety of interacting and overlapping factors like discretionary accounting practices or general stock market movements (P. J. Richard *et al.*, 2009). McWilliams, Siegel, and Teoh (1999) caution not to use share prices as a sole metric for CFP when investigating the CSP-CFP relation. Effects of specific plant or product-level information on value creation cannot be fully captured by simply looking at market-based

measures such as shareholder value. Therefore, the effects of CSP on operational CFP measures should be less likely to be drowned out by capital market ‘noise’ (Peloza, 2009, p. 1527). Empirical evidence for this argument was found in a meta-analysis on human resource factors (Combs *et al.*, 2006): operational CFP measures such as staff turnover yield higher correlations than other accounting- or market-based performance categories. Similar results were carried out in the environmental supply chain context: operational CFP measures such as reduced material and waste streams demonstrated slightly higher correlations to market-based measures (Golicic & Smith, 2013). Thus, we hypothesize:

H3b: *CSP has a stronger relation to operational CFP than to market-based CFP.*

Traditional CFP Categories

When considering the range of the traditional CFP measures, Cochran and Wood (1984) propose that accounting-based CFP such as return on assets may be the best proxy for measuring CFP – presuming that the accounting practices of firms are comparable. They more adequately reflect internal decision-making capabilities and managerial performance rather than external evaluations of investors. Empirical evidence supports this: accounting-based CFP measures have a stronger relation to CSP than market-based CFP measures (Margolis *et al.*, 2009; Orlitzky *et al.*, 2003; Wu, 2006).

Growth-based CFP – typically determined by sales, profit, or employee growth – is not directly affected by capital market developments. However, companies’ growth can be distorted by inorganic growth from acquisitions or divestments. Such activities usually have a relatively higher impact on growth measures (O’Shaughnessy & Flanagan, 1998) compared to changes in accounting figures (King, Dalton, Daily, & Covin, 2004). Accordingly, empirical research finds a weaker CFP relation for growth as a CFP measure compared to accounting-based CFP (Dixon-Fowler *et al.*, 2013; Unger *et al.*, 2011).

With regards to risk, primary research (Bansal & Clelland, 2004; Godfrey, Merrill, & Hansen, 2009; J. B. McGuire *et al.*, 1988) suggests that CSP reduces companies’ risk. Most of the empirical studies measure risk either as accounting-based risk (i.e. debt to assets) or market-based risk (i.e. beta-factor, total volatility). However, the observed correlations in meta-analyses are relatively low (e.g. Orlitzky & Benjamin, 2001). Summing up, these arguments bring us to the following hypothesis:

H3c: *Accounting-based CFP demonstrates a stronger relation to CSP than growth-, risk-, and market-based CFP.*

Mutual Funds

Several studies on mutual funds, financial indices, and virtual portfolios suggest that the positive CSP-CFP relation is weaker compared to studies using firm-specific CFP measures. To date, this result presents a puzzling outcome in CSP-CFP research. The first mutual funds studies appeared in the early 1990s. The initial findings of these early studies discovered that the performance of CSP-oriented mutual funds is not statistically different from the performance of conventional mutual funds (e.g. Hamilton, Jo, & Statman, 1993). However, ever since then, further investigations have been carried out in several primary and review studies (Bauer, Koedijk, & Otten, 2005; Capelle-Blancard & Monjon, 2014; Rathner, 2013; Revelli & Viviani, 2015), hardening the conclusion that the CSP-CFP link is weaker for mutual funds as compared to other CSP-CFP studies. In turn, we hypothesize:

H3d: *Studies on mutual funds demonstrate a weaker CSP-CFP relation compared to studies using other CFP measures.*

Moderating Factors

The literature discusses a range of conditions under which CSP and CFP influence each other (Aguinis & Glavas, 2012; McWilliams & Siegel, 2001). In addition to the most common factors – company size and the type of industry

(Aguinis & Glavas, 2012; Buehler & Shetty, 1974; Griffin & Mahon, 1997; Orlitzky, 2008) – several other moderating factors have been used in primary studies. These are visibility to the public and consumers (Fry, Keim, & Meiners, 1982; Jiang & Bansal, 2003), research and development investments (McWilliams & Siegel, 2000), economic conditions (Golicic & Smith, 2013), the level of diversification, advertising, government sales, consumer income, and stage in the industry as well as the firm life cycle (Aguinis & Glavas, 2012; Elsayed & Paton, 2009; McWilliams & Siegel, 2001). Since not all primary studies use the same bundle of control variables and/or moderating factors, it is difficult to compare such effects in meta-analyses. Thus, as meta-analysts have not adjusted for these factors in a comparable manner, we also cannot investigate these effects. Nonetheless, a few central moderating factors can be assessed at second-order level.

Effects over Time

It is not unusual that academic studies investigating a specific effect for the first time demonstrate impressive effect sizes which partly diminish in subsequent research studies (Trikalinos & Ioannidis, 2005). Griffin and Mahon (1997) have already called for research that looks at the CSP-CFP relation over time. To date, it is still an open question as to whether this is a time-independent relation. Scholars suggest that institutional investors increasingly focus on stakeholder issues (Borgers *et al.*, 2013; Derwall *et al.*, 2011). On account of learning effects and changes in investor values, these issues came into focus in recent years and subsequently the CSP-CFP relation – in this case measured as market-based CFP – diminished considerably since the mid-2000s. However, in meta-research on the CSP-CFP link, the evidence of a time dependency is not consistent. In their meta-analysis, Pavié and Filho (2008) show considerable effect size differences before and after 1998. Some discover no or only non-significant relations (Albertini, 2013; Endrikat *et al.*, 2014; Rubera & Kirca, 2012). Based on suggestions from primary studies and meta-analyses, we hypothesize about the moderating role of time:

H4a: *The CSP-CFP relation shrinks over time.*

Sample Size

Hedges (1981) demonstrates that effect sizes for small samples are biased upward. Evidence from cumulative meta-analyses in different fields supports this finding and cautions that meta-analyses with a small number of subjects may overestimate effects (Ioannidis, Trikalinos, Ntzani, & Contopoulos-Ioannidis, 2003; Trikalinos *et al.*, 2004). By using the number of included studies as a weighting factor, the calculation design of meta-analyses seeks to adjust for this circumstance. If the CSP-CFP relation was a statistical outlier, we would observe effect size differences depending on the study sample size. Initial observations of our sample based on the number of included primary studies in a meta-analysis and the corresponding meta-study overall effect size support such a conclusion. We therefore hypothesize:

H4b: *The CSP-CFP relation shrinks with the growing number of primary studies in the meta-sample.*

The Role of the Studies' Publishing Source

Orlitzky (2011) examines in a review of the 2003 data whether the empirical evidence on the CSP-CFP relation differs depending on the kind of publishing journal. The author finds that the average correlations in economics, finance, and accounting journals are less than half of the correlations published in social issues-oriented journals (SIM journals – social issues in management). The deviation is statistically significant. Allouche and Laroche (2005) reflect on their meta-analysis results by differentiating the publishing sources – business ethics journals, accounting journals, and management journals. They find no significant differences. In order to investigate the role of the publishing source for our sample, we follow a three-step approach. First, we differentiate between SIM and general management (GM) journals, similarly to Orlitzky (2011). Second, we classify the journals of our 25 meta-analyses according to the Thomson Reuters' Journal Citation Reports (JCR) classification standard [Social Sciences Citation Index (SSCI) subject categories]. Third, we take into consideration the JCR journal impact factor of the publishing journals for the meta-analyses in our sample. We expect that effect sizes will tend to be smaller in more rigorous journals (e.g. Aguinis, Beaty, Boik, & Pierce, 2005). We, thus, hypothesize:

H4c: *The CSP-CFP relation is distorted by the publication source; SIM and lower-impact journals report more positive outcomes.*

Methodological Quality

An analysis of meta-analyses in education research proposes that ‘there seems to be a particular tendency for smaller effect sizes to be associated with higher methodological quality’ (Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011, p. 14). However, the results were not significant. Based on nearly 200 meta-studies in management research, Aguinis, Dalton, Bosco, Pierce, and Dalton (2011) find that methodological choices have very little impact on the overall effect sizes. Lipsey and Wilson (1993) reveal for 302 studies in psychology that effect sizes in meta-studies are not inflated by the inclusion of methodological weak studies. To create generalizable conclusions for the CSP-CFP debate, we build on Aguinis *et al.* (2011) and apply a methodological strength assessment. We hypothesize:

H4d: *The CSP-CFP relation is independent of methodological choices*

Methods and Data

The first comprehensive summary of meta-analyses was published at the beginning of the 1990s, aggregating the work of 302 meta-analyses (Lipsey & Wilson, 1993). Ever since then, various examples of how to aggregate existing research summaries have emerged (Ahn, Ames, & Myers, 2012; Aytug, Rothstein, Zhou, & Kern, 2012; Cooper & Koenka, 2012; Hattie, 2009; Peterson, 2001; D. Richard, Bond, & Stokes-Zoota, 2003; Tamim *et al.*, 2011; Wilson & Lipsey, 2001). The main challenges of second-order meta-analyses are the precise calculation of the second-order average effect and estimation of its sampling error. Hunter and Schmidt (2004) and Borenstein, Hedges, Higgins, and Rothstein (2009) already address these challenges, but only recently have sufficient statistical methods been provided.

This study uses Schmidt and Oh’s (2013) method for second-order meta-analysis. A second-order meta-analysis combines a number of independent but methodologically comparable first-order meta-analyses. It enables better generalization across contexts with a tremendously larger sample size and enhanced meta-analytic moderating factor analyses. Conclusions can be derived not just from a fraction of available primary studies, but presumably with a sample close to a complete set of studies. The Schmidt and Oh (2013) second-order meta-analysis is an advancement of the psychometric first-order predecessor – the Hunter-Schmidt (2004), approach, which is used by more than 80% of meta-analyses in management research (Aguinis *et al.*, 2011).

Schmidt and Oh’s (2013) method has two important advantages compared to other second-level meta-analytic approaches. First, their method determines how much variance across a set of comparable first-order meta-analyses is due to second-order sampling error. This can considerably reduce the remaining sampling error variance of first-order meta-analyses and allows estimation of the true (non-artifactual) variance across these mean effect sizes. Second, the analysis enables the computation of reliability estimates for the first-order meta-analyses and, thus, allows us to produce a more accurate estimation of the true mean effect size in each first-order meta-analysis.

Schmidt and Oh (2013) present three approaches for second-order meta-analysis: bare-bones second-order meta-analysis; psychometric second-order meta-analysis with artifact distribution of first-order meta-analysis estimates; and psychometric second-order meta-analysis with individual corrected first-order meta-analysis estimates. Because of the high number of first-order meta-analyses in our sample using artifact distribution correction without correcting primary studies individually, we opted for the first two approaches to calculate our results. Both approaches vary in terms of the input data (uncorrected or attenuated correlations vis-à-vis corrected or disattenuated correlations), the weighting scheme of the first-order meta-analyses, and the calculation of the second-order mean effect size.¹

¹See Table 1 in Schmidt and Oh (2013, p. 210) for a technical summary of the three approaches.

Like the first-order meta-analytical predecessor methodology (Hunter *et al.*, 1982; Hunter & Schmidt, 2004), second-order meta-analysis is also a fully random effect model.

The Search Process

We included a CSP-CFP meta-analysis in our sample insofar the applied CSP definition of meta-analysis matched the CSP understanding expressed by Wood (2010) and Clarkson (1995). Wood (2010, p. 64) argued for ‘CSP as an overarching, multi-dimensional construct’ where vast literature on organizational culture, managerial decision-making, or employee relations practices ‘waits to be brought into the field of CSP’. Also, Clarkson (1995) outlined a diverse set of social and stakeholder issues. He included, for example, employee-related aspects like compensation/rewards, training and development, employment equity, and employee communication. We follow their line of reasoning and have also included meta-analyses beyond a narrow CSP focus. These are, for instance, CEP, ethical corporate governance, and broader employee-related aspects. We did not differentiate as to whether the motives for firm CSP were altruistic or strategic (Baron, 2001; J. W. McGuire, 1969; McWilliams, Siegel, & Wright, 2006). In terms of CFP, the studies needed to consider at least one of the following categories: accounting-based performance, market-based performance, operational performance, perceptual performance, growth metrics, risk, or the performance of CSP mutual funds/indices. In cases where the entire meta-analysis was beyond this scope, we extracted only the relevant parts of the whole analysis.² Meta-analyses with probit- or logit-correlations were excluded, as transformations to the r-family of correlations result in effect sizes with too large sampling errors. Non-standardized effects, which could not be classified as CSP-CFP sub-effects, were also disregarded for this second-order meta-analysis. Our cut-off date for study inclusion was December 2014. All studies needed to be available in electronic format.

Initially, we applied ancestry research, using widely cited meta- and review studies to extract their cited predecessor studies. The sample was subsequently extended with a forward search in Google Scholar looking for potentially relevant cross-references of so far unknown studies. After that, we performed searches in relevant databases and publisher sites: Academy of Management journals, ABI/Inform, Ebsco, Emerald, Google Scholar, Oxford Journals, Sage, Science Direct, Springer Link, Web of Science, and Wiley, and also searched for non-published manuscripts on Econbiz, National Bureau of Economic Research (NBER), Research Papers in Economics (RePEc), and Social Science Research Network (SSRN). We used several search combinations and variations of the term ‘Corporate Social Performance’ including (corporate) sustainability, (corporate) social responsibility, (corporate) governance, (corporate) environmental performance – all in relation to (corporate) financial performance. Of all database queries, we further processed the first 100 hits sorted by relevance. The raw data set of several thousand studies was first reduced for overlapping search hits, and second searched for keywords in the remaining results. We searched for meta, review, literature, overview, analysis, study/ies, and examination. This yielded an overall raw sample of 149 working papers and studies which were analyzed in more detail by examining the abstract or full paper. Seventy-six studies were dropped in the next step, because they either did not contain a relevant CSP-CFP categorization or were single-study designs. The remaining studies were narrowed to the final sample of 25 studies by the exclusion of: (1) 47 narrative reviews and vote-count studies; and (2) meta-studies with non-transferable effect sizes and different paper versions. If different versions of meta-studies with comparable data sets were identified, the latest version remained in our sample. Our search was open to meta-analyses in all asset classes, but it was not possible to locate meta-studies that analyzed other objects than equities. Thus, published (primary) research of the CSP effect in corporate bonds, real estate, and private equity is not considered, as we could not yet identify dedicated meta-analyses.

Sample

Our total sample consists of 25 meta-analyses which combined yield a gross number (n_m) of 1,902 primary studies. It aggregates 4,507 separate effects (k_m) and nearly a million observations ($N_m = 992,239$) (Table 1). Apart from the meta-analytical summary effects, we obtained 129 CSP-CFP sub-effects of first-order meta-analyses, combining in total 2,387,215 observations. Besides the summary and sub-effects at the meta-level, we extracted all provided data at primary study level. Adjusted for overlapping studies and primary studies which either were not made

²Further details are provided in online Supplementary Appendix SA.

References	Transparency of primary studies				Statistics of summary effects				
	n_m	N_m	yes/no	avg. overlap	statistics	\bar{T}_i	$\hat{\sigma}_{\bar{T}_i}^2$	\hat{P}_i	$\hat{\sigma}_{\hat{P}_i}^2$
Frooman (1997)	22	2,161	X	2.8%	n, t-stats	0.312(d)	0.044(e)	0.436(a)	0.086(d)
Orlitzky and Benjamin (2001)	18	6,186	X	9.9%	N, r, reliab.	0.149	0.048	0.209	0.070
Orlitzky (2001)	20	6,889	X	10.9%	N, r, reliab.	0.061	0.020	0.105	0.039
Orlitzky <i>et al.</i> (2003)	62	33,878	X	13.6%	N, r, reliab.	0.184	0.065	0.365	0.190
Allouche and Laroche (2005)	79	57,409	X	14.1%	N, r, t-stats	0.143	0.001(c)	0.199(d)	0.003(d)
Combs <i>et al.</i> (2006)	90	19,319	X	0.7%	n.a.	0.150	0.013	0.200	0.050
Wu (2006)	120	21,933	X	9.0%	n.a.	0.166	0.008(e)	0.231(d)	0.016(d)
Rosenbusch, Bausch, and Calander (2007)	62	21,742	no	—	n.a.	0.190	0.016(e)	0.264(d)	0.031(d)
Darnall and Sides (2008)	9	30,000	X	0.6%	r	0.077	0.008(c)	0.107(d)	0.016(d)
Pavie and Filho (2008)	112	170,737	X	1.7%	N, r, reliab.	0.083	0.029	0.180	0.117
van Wijk, Jansen, and Lyles (2008)	28	4,627	X	0.4%	n.a.	0.190	0.025	0.220	0.049(d)
Margolis <i>et al.</i> (2009)	214	38,483	X	17.5%	n.a.	0.133	0.005(e)	0.185(d)	0.009(d)
Vishwanathan (2010)	189	n.a.	no	—	n.a.	0.070	0.011(c)	0.097(d)	0.022(d)
Crook <i>et al.</i> (2011)	66	12,163	X	2.3%	N, r	0.170	0.010	0.210	0.011(c)
Rosenbusch, Brinckmann, and Bausch (2011)	46	21,270	X	1.0%	N, r	0.133	0.013(c)	0.185(d)	0.025(d)
Unger <i>et al.</i> (2011)	70	24,733	X	3.1%	N	0.076	0.005	0.098	0.002(c)
Rubera and Kirca (2012)	153	33,544	no	—	n.a.	0.146	0.005(b)	0.203(d)	0.010(d)
Albertini (2013)	52	62,943	X	7.2%	N, r	0.090	0.000(c)	0.125(d)	0.000(d)
Dixon-Fowler <i>et al.</i> (2013)	39	22,869	X	8.0%	n.a.	0.062	0.003(b)	0.086(d)	0.007(d)
Golicic and Smith (2013)	31	15,160	X	1.2%	N, r	0.305	0.004(d)	0.294	0.008(c)
Mayer-Haug, Read, Brinckmann, Dew, and Grichnik (2013)	175	50,045	X	4.0%	N	0.044(b)	0.023(d)	0.055	0.036(c)
Endrikat <i>et al.</i> (2014)	148	201,511	X	6.7%	N	0.082(b)	0.012(d)	0.102	0.018(c)
Stam, Arzalian, and Elfring (2014)	43	13,263	X	1.8%	N, r	0.157	0.009(d)	0.211	0.018(c)
del Mar Miras-Rodríguez, Carrasco-Gallego, and Escobar-Pérez (2015)	91	31,878	X	1.5%	N	0.067	0.046(c)	0.094(d)	0.090(d)
Revelli and Viviani (2015)	80	89,496	X	1.4%	d	-0.003(a)	0.068(a)	-0.003(d)	0.133(d)
Total / Simple Average	1,902	992,239		5.4%		0.129	0.020	0.178	0.042

Table 1. Overview of the meta-studies in the sample.
 Note: Deviations for the stated numbers for primary studies n_m in comparison to figures in the first-level meta-analyses result from (1) elimination of studies that are counted multiple times, (2) addition of studies if a higher number of unique studies were outlined in tables, and (3) extracting only the relevant CSP-CFP sub-sample for our analysis if not the total sample was used. The indices (a) to (e) indicate the source of the effect size and variance: (a) transformed from d / g in r ; (b) transformed from given standard errors to r ; (c) variance derived from provided credibility or confidence intervals; (d) derived with either an individual or the artifact attenuation factor of 0.72 ($\hat{\sigma}_{\bar{T}_i}^2 = \hat{\sigma}_{\bar{T}_i}^2 * a^2$) and; (e) variance estimated based on study sample size n_i and overall n -weighted attenuated effect size r with formula $\hat{\sigma}_{\bar{T}_i}^2 = \frac{r_i - 1}{n_i - 1}$. See text for abbreviations.

transparent by the meta-analyses' authors ($n = 404$) or needed manual corrections ($n = 38$), the final study sample comprises 1214 (n) unique studies. Manual corrections were necessary for author typing errors, different years (i.e. published vs unpublished version), or adding or changing 'a' or 'b' when two different studies of the same author team appeared in the same year. When retrieving the primary study information from the meta-studies, we coded it in a unique form as author₁, author₂, ..., author i (year). For 45% of studies ($n_{(i)} = 551$ primary studies), the first-order meta-analyses provided data about effect sizes. On the basis of this information, it was possible to calculate a conventional first-order meta-analysis (so-called second-order omnibus meta-analysis).

Coding

Besides the basic identifier information: (1) we retrieved a variety of methodological information. This included the meta-analytical approach, the family of reported effect sizes, if and how corrections for sampling and measurement error were taken, and, finally which attenuation factors were used. (2) We collected the granularity of primary studies (transparency level of statistics, all effect sizes, N , k , and t -test). (3) Finally, we gathered all meta-analytical results for summary and sub-effects (effect sizes, attenuated and disattenuated variances, standard errors, credibility and confidence intervals, reliability information of dependent and independent variables, the overall reliability estimate, and results of the homogeneity test, significance tests, and Fail-Safe N analysis). If results for both fixed and random effects were reported, we selected the random effect. All transparent data in primary studies which were also included in a first-order meta-analysis were extracted (e.g. the study identifier and results like study size, effect size(s), and reliabilities). In some cases, rudimentary interpretation was needed. For these cases, two independent coders reviewed the context. The original inter-rater reliability was above 0.9.

Independence Considerations

An important prerequisite for conducting a second-order meta-analysis is statistical independence of individual first-order meta-analyses. This would require that the primary studies or samples used in one first-order meta-analysis not be included in any of the other analyses (Schmidt & Oh, 2013). In fact, it is a typical case for second-order meta-analyses that this assumption cannot completely be met. Scholars therefore suggest that minimizing the lack of independence might be the best that can be expected (Cooper & Koenka, 2012; Schmidt & Oh, 2013). Various simulation analyses support this approach and find negligible effects of statistical independence violations (Bijmolt & Pieters, 2001; Romano & Kromrey, 2008; Tracz, Elmore, & Pohlmann, 1992).

Our approach for minimizing data dependency applied three steps: meta-analyses selection; average overlap assessment; and the elimination of overlapping studies as a robustness check. First, we selected only the most recent version of a study. We considered multiple contributions from single authors only if they had different research topics or reported different correlation coefficients for the same primary study (Orlitzky, 2001; Orlitzky *et al.*, 2003; Orlitzky & Benjamin, 2001). Second, we calculated for every meta-analysis with transparent primary studies the pair-wise overlap to each other. We determined an average overlap of 5.4% of one single meta-analysis to the rest with 78% of pairs having less than 5% overlap. Third, Wilson and Lipsey (2001) reduce disproportionate dependency in second-order meta-analysis by eliminating studies that have more than 25% primary studies in common with other meta-analyses. Thus, we eliminated all meta-analyses with more than 25% overlap to at least one other study. The results of an additional assessment to rule out distorting effects of a potential independence violation are reported in Table 6. We conclude that the level of average non-independence for this study appears small and we have no indication that it could distort effect sizes and variances significantly.

Analysis and Statistical Models

We calculated summary effects in three alternative ways and also for various subgroups and moderating factors. First, we conducted a second-order omnibus meta-analysis with all primary studies in our meta-analyses sample which contained transparent data for effect sizes. Technically, this is a traditional psychometric first-order

meta-analysis (Hunter *et al.*, 1982; Schmidt & Hunter, 2015). So-called omnibus meta-analyses have the advantage of established calculation procedures from first-order meta-analysis. Moreover, such analyses yield the same grand mean estimate, provided that all else is equal. But they also suffer from shortcomings like typically scarce data transparency at primary study level and, most important, it is not possible to estimate second-order sampling error variance and correct for it (Schmidt & Oh, 2013). This omnibus meta-analysis, to our knowledge, constitutes the largest meta-analysis in the field hitherto, based on the number of included primary studies. It combines $n_{(i)} = 551$ studies with $N_{(i)} = 152,437$ observations retrieved from previous meta-analyses.³ We utilized this first-order meta-analysis to gain a more granular understanding of the CSP-CFP relation and its statistical variation at primary study level.

Second, we calculated a second-order meta-analysis combining all meta-analytical summary effects of the 25 meta-analyses in our sample ($m = 25$).

Third, we aggregated all relevant sub-effects of these meta-analyses ($m = 129$) in various subgroups in order to derive more granular conclusions about the CSP-CFP relation. These sub-effects and their reliability measures provided the foundation for the calculation and correction of a second-order sampling error and the subsequent moderating factor analysis. Moreover, as part of the robustness checks, we analyzed the impact on results when changing from best-effort variances to simple variance estimates. Further documentation on the determination of effect sizes and variances, the correction for measurement error, and the assessment of heterogeneity is provided in online Supplementary Appendix SA.

Adjusting for the Second-Order Sampling Error

The second-order sampling error is comparable to adjustments for first-order measurement error variance in psychometric meta-analysis. As measurement error is the random deviation of an observed score from a true score, so is the random deviation of first-order meta-effect sizes from its actual second-order average, the second-order sampling error. Thus, the function is comparable to measurement error (Schmidt & Oh, 2013):

$$\widehat{p}_{ir} = r_{pp} (\widehat{p}_i - \widehat{p}) + \widehat{p} \quad (1)$$

where \widehat{p}_{ir} is the second-order sampling error adjusted for the i -th first-order meta-analytic estimate and unreliability in the first-order meta-analytic results, r_{pp} is the reliability of i first-order meta-analytic values, \widehat{p}_i is the i -th first-order meta-analytic estimate, and \widehat{p} is the second-order grand mean meta-analytic estimate. The reliability coefficient r_{pp} is 1-ProVar, which is calculated as the ratio of $\widehat{\sigma}_{\widehat{p}}^2 / S_{\widehat{p}}^2$ where $\widehat{\sigma}_{\widehat{p}}^2$ is the estimate of the actual (non-artifactual) variance across the m -meta-analyses of the population mean-disattenuated correlations \widehat{p} . This is the variance that remains after the variance for the second-order sampling error $E\left(S_{\widehat{p}_i}^2\right)$ has been subtracted from the variance of mean-disattenuated correlations $S_{\widehat{p}}^2$ across the m -meta-analyses (Schmidt & Oh, 2013).

Adjustments for second-order sampling error are useful to reduce the remaining sampling error variance of first-order meta-analyses and therefore produce more accurate estimations of the true mean effect size and (non-artifactual) variance in first-order meta-analyses. These figures are displayed in columns 6–13 of Table 4.

Moderating Factors

For determining potential moderating factors, the data were split into two or, if necessary, into multiple subgroups. We calculated z -values to identify whether effect size differences among groups are significant. For assessing the potential time dependence of the CSP-CFP relation, we analyzed the effect sizes in dependence of the meta-study publication year and the average publication year of the primary study in each meta-analysis. To rule out a distortion

³The index (i) for n and other statistics symbolizes first-order level (extracted effects of primary studies from meta-analyses) in comparison to figures without this index which are second-order level.

of results from changing meta-analytic methods and correction factors over time, the relation was as well assessed on first-order level and in dependency of the publication year.

The number of primary studies in each meta-analysis and its repercussions for the correlation level were analyzed for evaluating the role of sample sizes. The influence of the publishing source was compared for SIM journals versus General Management (GM) journals based on Orlitzky's (2011) categorization. Since definitions regarding what can be considered as a SIM or GM journal differ (AoM SIM Division, 2012), all publishing journals of the 25 meta-analyses were classified according to the Thomson Reuters' JCR classification standard. The SSCI subject categories present in our sample (business, environmental studies, ethics, management, political science, and applied psychology) were used to alternatively differentiate the sample of meta-analyses. If a journal was listed in more than one SSCI category, the one where it ranks highest was selected. Furthermore, the JCR 5-Year Journal Impact Factors (2014), the annual citation frequency of the meta-analyses, and the deviation between working papers and published studies were evaluated.

Finally, effect size dependence on methodological quality was evaluated. The methodology quality assessment builds on Aguinis *et al.* (2011) and transferred those parts of Aguinis *et al.*'s 21 parameters which could be expressed as dichotomous and qualitative. All parameters which were applied multiple times (e.g. range restrictions) are summarized to one overarching parameter. For each of the parameter identifiable in a meta-study, the meta-study received a score of 1 (present) or 0 (not present). The condensed set contained 11 parameters and therefore yielded a maximum methodological quality score of 11. The total scores for the set of 25 meta-analyses range between 2 and 9.

Addressing Publication Bias

A general concern of any meta-analysis is that research yielding no significant results remains unpublished (e.g. Dickersin, 2005). Rosenthal (1979) labelled this potential limitation as the file drawer problem, with reference to the many files that remain with the investigator and are not published by journals. We addressed this potential drawback by calculating the Fail-Safe N (or file drawer) statistic, a display of the Funnel Plot, and the inclusion of unpublished literature. Fail-Safe N states the number of (future) studies with null results, until the effect size loses its significance level (Rosenthal, 1979). Rosenthal suggested a tolerance level of $5n+10$ for Fail-Safe N to be considered unlikely. The statistic was calculated for all summary and sub-effects. For a visual investigation of publication bias we included as well a funnel plot (Light & Pillemer, 1984; Sutton, 2009). It scatters the study size of a meta-analysis against the effect size. In our case the funnel plot displays all $m = 129$ sub-effects (number of effects k) against its disattenuated effect size p (Figure 1). The effects are symmetrically distributed around the underlying true effect, with larger variability for smaller studies (and vice versa). As such, we can exclude biases in the data. Furthermore, we applied the Egger test for publication bias, which tests for asymmetry of the funnel plot (Egger, Smith, Schneider, & Minder, 1997). Moreover, we separated attenuated and disattenuated effects from unpublished studies and subgroup the results against published studies (Schmidt & Hunter, 2015). The proportion of these working papers within the total results was notably high. They contributed 28 of the 129 sub-effects. All checks revealed that the results are very likely not be affected by publication bias (Figure 1; Table 5).

Robustness Checks

The empirical outcomes of this analysis can be considered as fairly robust for a number of reasons. First, we analyzed the impact of switching from our best-effort variances to simple variance estimates (calculation according to Equation 6 in Online Supplement). The results in Table 6 show that our best-effort variances are conservative estimates, which produce lower effect sizes and variances, in particular for disattenuated results. Second, we performed a sensitivity analysis (Aguinis, Gottfredson, & Joo, 2013; Greenhouse & Iyengar, 2009) and recalculated all effects when leaving out the top 5% and bottom 5%, each for effect size and effect size variances. Third, to control for a potential bias through perceptual measures of CFP and CSP, the effects were recalculated by leaving out the complete sample of perceptual CFP and reputational CSP. The results of all three robustness checks indicate that the findings are not distorted by outliers (Table 6).

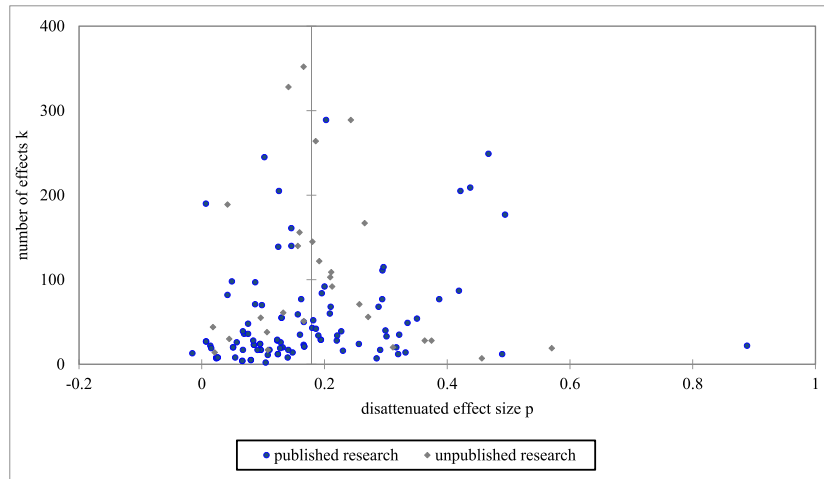


Figure 1. Funnel plot of meta-analyses' sub-effects ($m = 129$) against disattenuated effect size. *Note:* Figure 1 scatters the study size of the 129 meta-analysis sub-effects (number of effects k) against its disattenuated effect size p . Effects of published and unpublished studies are displayed separately. The intersection of the axes was moved to the simple average of disattenuated effect sizes (0.179). The effects are relatively symmetrically distributed around the average effect, with larger variability for smaller studies (and vice versa). The larger the number of effects k , the closer the effect sizes are expected to be around the population mean. Effect sizes of unpublished research do not indicate any suppression of unfavourable studies. Apart from a very few outliers, which are additionally accounted for in the Robustness Checks section, the sample is not distorted by publication bias. This visually derived conclusion is supported by the Egger test (Egger *et al.*, 1997) with a calculated p -value for bias of 0.489. [Colour figure can be viewed at wileyonlinelibrary.com]

Results

The General CSP-CFP Relation

We find strong support for Hypothesis 1a. First, the second-order omnibus meta-analysis, which is technically a first-order meta-analysis with $n_{(1)} = 551$ primary studies, reveals a highly significant (first-order) effect size of $\bar{r}_{(1)} = 0.119$ (Table 2). This number presents the weighted average correlation across all primary studies adjusted for sampling error. When we additionally adjust for meta-analytic artifacts (measurement error for the CSP and CFP figures), we find a disattenuated correlation of $\widehat{\bar{p}}_{(1)} = 0.169$. The average effect is highly significant and different from zero. The 95% confidence intervals for this weighted average attenuated correlation range from 0.105 to 0.134 (0.150 to 0.189 for the disattenuated). The 95% credibility interval ranges from -0.206 to 0.444 (-0.284 to 0.624 for the disattenuated). The corresponding variance $\widehat{\sigma}_{\widehat{\bar{p}}_{(1)}}^2$ of 0.054 (Table 2) is fairly comparable to the simple average of

$n_{(1)}$	$N_{(1)}$	$\bar{r}_{(1)}$	$S^2_{\bar{r}_{(1)}}$	$\widehat{\sigma}_{\bar{r}_{(1)}}^2$	$StdE_{(1)}$	$CI_{L(1)}^{95\%}$	$CI_{U(1)}^{95\%}$	$CrI_{L(1)}^{95\%}$	$CrI_{U(1)}^{95\%}$	$1-ProVar_{(1)}$	$I^2_{(1)}$	$Q_{(1)}$	Fail-Safe $N_{(1)}$
551	152,437	0.119***	0.032	0.027	0.008	0.105	0.134	-0.206	0.444	0.89	0.88	4,733***	164,294
$n_{(1)}$	$N_{(1)}$	$\widehat{\bar{p}}_{(1)}$	$S^2_{\widehat{\bar{p}}_{(1)}}$	$\widehat{\sigma}_{\widehat{\bar{p}}_{(1)}}^2$	$StdE_{(1)}$	$CI_{L(1)}^{95\%}$	$CI_{U(1)}^{95\%}$	$CrI_{L(1)}^{95\%}$	$CrI_{U(1)}^{95\%}$	$1-ProVar_{(1)}$	$I^2_{(1)}$	$Q_{(1)}$	Fail-Safe $N_{(1)}$
551	152,437	0.169***	0.060	0.054	0.010	0.150	0.189	-0.284	0.624	0.89	0.94	9,288***	169,266

Table 2. First-order meta-analytical results (attenuated bare-bones results and psychometric results with artifact correction). *Note:* Table 2 outlines the first-order meta-analytical results based on extracted statistics for primary studies from $m = 25$ meta-studies (second-order omnibus meta-analysis). The additional index (1) is introduced to differentiate these first-order meta-analytical results from second-order results. The significance values are:

* $p < 0.10$;
 ** $p < 0.05$ and
 *** $p < 0.01$.

disattenuated correlations in the meta-analyses in our sample of 0.042 (Table 1). The effects are very heterogeneous [$Q_{(1)} = 4733$ to 9288 and $I^2_{(1)} = 0.88$ to 0.94].

Second, the aggregation of all summary effects of the meta-analyses in our sample ($m = 25$) uncovers effect sizes that are very similar to those in the sample of primary studies. We determine an attenuated correlation of $\bar{r} = 0.108$ and a disattenuated correlation of $\hat{\bar{p}} = 0.150$ (Table 3).

Third, the correlations change marginally when the sample of 129 sub-effects from the meta-analyses is aggregated ($\bar{r} = 0.110$, $\hat{\bar{p}} = 0.157$). But unlike effect sizes, (true) variances at second-order level are considerably lower compared to those in the omnibus analysis of primary studies. Second-level variances are only in the range of $\hat{\sigma}^2_{\bar{r}} = 0.001$ to $\hat{\sigma}^2_{\hat{\bar{p}}} = 0.007$ (Table 3). The 95% credibility interval for disattenuated sub-effects ranges from -0.001 to 0.316. Moreover, in our second-order results, we cannot detect the large heterogeneity of effects anymore that was present in the primary studies. Q and I^2 indicate homogeneous relations ($Q = 11.4$ to 75.5; $I^2 = 0.00$; Table 3). The third and preferred measure of heterogeneity for psychometric meta-analysis (1-ProVar) still indicates heterogeneity (0.96 to 0.97) – albeit in very small variances, which have low statistical power.

We conclude that all CSP-CFP summary effects, be it first or second order, are highly significant positive. Second-order level variances are furthermore very small with credibility intervals that only by a very thin margin include zero and with effect sizes that are relatively homogeneous.

With respect to Hypothesis 1b, the difference between prior CSP to subsequent CFP ($\bar{r}_i = 0.090$) vs prior CFP to subsequent CSP ($\bar{r}_i = 0.098$) is small and non-significant (Table 4). We apply two layers of corrections to account for other artifacts besides first-order sampling error. Based on the first-order correction for other artifacts, the relation changes slightly ($\hat{\bar{p}}_i = 0.128$ vs $\hat{\bar{p}}_i = 0.141$). The additional correction for second-order sampling error reveals that the real grand mean correlation is marginally different ($\hat{\bar{p}}_{ir} = 0.130$ vs $\hat{\bar{p}}_{ir} = 0.143$). The difference between both groups fails, with a wide margin, any significance level. We, thus, can rule out that the slack resources hypothesis or the good management hypothesis yields significantly higher results. However, we find the strongest correlation for concurrent CSP and CFP, which supports Hypothesis 1b that there is a virtuous circle between CSP and CFP. The mean correlation for concurrent CSP and CFP, corrected for second-order sampling error, is $\hat{\bar{p}}_{ir} = 0.159$, the highest of the group with the lowest variance.

	<i>m</i>	<i>k_m</i>	<i>N_m</i>	\bar{r}	$S^2_{\bar{r}}$	$\hat{\sigma}^2_{\bar{r}}$	StdE	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	1-ProVar	<i>I</i> ²	<i>Q</i>	Fail-Safe <i>N</i>
	<i>m</i>	<i>k_m</i>	<i>N_m</i>	$\hat{\bar{p}}$	$S^2_{\hat{\bar{p}}}$	$\hat{\sigma}^2_{\hat{\bar{p}}}$	StdE	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	1-ProVar	<i>I</i> ²	<i>Q</i>	Fail-Safe <i>N</i>
a)	25	1,902	992,239	0.108***	0.001	0.001	0.007	0.094	0.123	0.035	0.182	0.97	0.00	18.1	717
a)	129	8,337	2,387,215	0.110***	0.004	0.004	0.005	0.099	0.120	-0.009	0.229	0.96	0.00	73.2	1,681
b)	25	1,902	992,239	0.150***	0.002	0.002	0.009	0.132	0.169	0.058	0.242	0.96	0.00	11.4	987
b)	129	8,337	2,387,215	0.157***	0.007	0.007	0.007	0.143	0.171	-0.001	0.316	0.96	0.00	75.5	1,861

Table 3. Second-order meta-analytical results (summary effects $m = 25$; meta sub-effects $m = 129$).
 Note: Table 3 depicts the second-order meta-analytical results: a) for the bare bones second-order meta-analysis; and b) for the psychometric second-order meta-results corrected with the artifact distribution factor. *m* stands for the number of included meta-analytical effects ($m = 25$ is the highest level of reported meta-analytical summary-effects, $m = 129$ are extracted meta-analytical sub-effects); *k_m* is the number of studies at primary level; *N_m* states the number of the overall sample size in the *m* metas; \bar{r} is the variance-weighted attenuated average second-order effect size; $\hat{\bar{p}}$ is the variance-weighted disattenuated average second-order effect size; S^2 is the corresponding sample-size weighted observed variance for \bar{r} and $\hat{\bar{p}}$; $\hat{\sigma}^2$ is the estimated true variance for \bar{r} and $\hat{\bar{p}}$; StdE is the standard error; $CI_{95\%}$ is the 95% confidence interval and $CrI_{95\%}$ is the 95% credibility interval, in which L indicates the lower and U the upper bounds; 1-ProVar describes the reliability of the first-order meta-analytical correlations and indicates the proportion of (true) variance which is not due to first- and second-order artifacts; *I*² and *Q* are the corresponding heterogeneity measures. The significance values are:

* $p < 0.10$;
 ** $p < 0.05$; and
 *** $p < 0.01$.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
n_m	k_m	N_m	$\bar{\tau}_i$	$S_{\tau_i}^2$	$\widehat{\bar{p}}_i$	$S_{e_{\widehat{p}_i}}^2$	$\widehat{\bar{p}}$	$E(S_{e_{\widehat{p}_i}}^2)$	$S_{\widehat{p}}^2$	$\widehat{\sigma}_{\widehat{p}}^2$	$1 - ProVar$	$\widehat{\bar{p}}_{ir}$	$z\text{-value}$ $\widehat{\bar{p}}_1 \text{ vs. } \widehat{\bar{p}}$
Causality													
CSP = CFP	8	845	248,125	0.103***	0.0014	0.160***	0.148	0.0007	0.0055	0.0048	0.88	0.159	0.51
CFP --> CSP	8	261	55,510	0.098***	0.0018	0.141***	0.159	0.0001	0.0035	0.0031	0.89	0.143	0.23
CSP --> CFP	8	325	141,027	0.090***	0.0032	0.128***	0.159	0.0001	0.0068	0.0049	0.71	0.130	0.59
CSP focus													
CSP (ex CEP)	8	1,210	273,649	0.127***	0.0012	0.181***	0.225	0.0008	0.0022	0.0020	0.93	0.179	1.05
CEP	11	1,171	436,350	0.106***	0.0028	0.151***	0.225	0.0008	0.0040	0.0039	0.98	0.151	0.34
CSP dimension													
Reputation	6	386	42,837	0.221***	0.0004	0.318***	0.148	0.0003	0.0036	0.0029	0.80	0.299	2.66***
Philanthropy	6	344	18,953	0.175***	0.0133	0.246***	0.148	0.0003	0.0262	0.0255	0.97	0.245	0.29
Audits, processes, policies	5	271	52,695	0.098***	0.0009	0.137***	0.148	0.0003	0.0017	0.0004	0.25	0.203	3.09***
Disclosure	5	408	129,925	0.067***	0.0003	0.119***	0.148	0.0003	0.0022	0.0013	0.58	0.163	3.39***
CFP category													
Perceptual	5	273	52,688	0.256***	0.0068	0.380***	0.148	0.0003	0.0079	0.0076	0.96	0.359	5.45***
Operational	5	168	47,328	0.170***	0.0051	0.222***	0.148	0.0003	0.0093	0.0085	0.91	0.216	2.80***
Accounting	23	1,265	381,503	0.115***	0.0035	0.161***	0.148	0.0003	0.0032	0.0029	0.92	0.161	0.67
Growth	8	223	56,280	0.092***	0.0011	0.124***	0.148	0.0003	0.0065	0.0062	0.95	0.130	1.32
Market	19	809	252,601	0.080***	0.0024	0.115***	0.148	0.0003	0.0041	0.0039	0.94	0.117	1.81*
Risk	2	158	104,977	0.045**	0.0010	0.062**	0.148	0.0003	0.0020	0.0017	0.87	0.074	2.75***
Mutual funds	2	220	92,767	0.011	0.0001	0.016	0.148	0.0003	0.0003	0.0000	0.00	0.148	8.37***

Table 4. Second-order meta-analytical results for subgroups.

Note: Table 4 describes second-order meta-analytical results for various subgroups. n_m is the number of included meta study sub-effects; k_m is the number of studies in the n_m ; N_m states the number of the overall sample size in the n_m ; $\bar{\tau}_i$ is the variance-weighted attenuated average second-order effect size; $S_{\tau_i}^2$ is the corresponding sample size-weighted observed variance for $\bar{\tau}_i$; $\widehat{\bar{p}}_i$ is the variance-weighted disattenuated average second-order effect size with the corresponding second-order sampling variance $S_{e_{\widehat{p}_i}}^2$; the expected

(average) second-order sampling error for the group is denoted as $E(S_{e_{\widehat{p}_i}}^2)$ and $\widehat{\bar{p}}$ is the average second-order grand mean effect size for the group; $S_{\widehat{p}}^2$ symbolizes the observed variance for $\widehat{\bar{p}}$ and $\widehat{\sigma}_{\widehat{p}}^2$ is the estimated true variance for $\widehat{\bar{p}}$; 1-ProVar describes the reliability of the first-order meta-analytical correlations and indicates the proportion of (true) variance which is not due to first- and second-order artifacts; $\widehat{\bar{p}}_{ir}$ symbolizes the overall second-order meta-analytical effect size estimate corrected for second-order sampling error. In column 14, the respective z-values and significance values of the difference between $\widehat{\bar{p}}_1$ and the group average $\widehat{\bar{p}}$ are displayed. See text for abbreviations. The significance values are:

* $p < 0.10$;
 ** $p < 0.05$; and
 *** $p < 0.01$.

The Relation of Various CSP Dimensions and CFP

In line with Hypothesis 2a, we find that studies with social CSP reveal a higher relation ($\hat{\rho}_i = 0.181$) vs studies focusing on environmental CSP ($\hat{\rho}_i = 0.151$) (Table 4). However, this difference is small and fails statistical significance (p-value of the difference is 0.388). Thus, Hypothesis 2a is rejected. Opposite to a common notion in the literature, our findings propose that good CSP pays, whether social or environmental related.

Within CSP dimensions, CSP reputation yields the highest relation ($\hat{\rho}_i = 0.318$). The difference to the group average of CSP measures of $\hat{\rho} = 0.225$ is significantly higher at the 99% level (z-value 2.66). However, when we adjust for the 0.80 reliability factor (Table 4, column 12), CSP reputation is marginally adjusted towards $\hat{\rho}_{ir} = 0.299$ to account for the second-order sampling error. Thus, we confirm Hypothesis 2b: reputational aspects have the strongest relation to CFP within the CSP dimensions.

CSP disclosure yields significantly lower correlations ($\hat{\rho}_i = 0.119$) compared to the group average ($\hat{\rho} = 0.225$) at the 99% level (Table 4). On account of the high level of variance artifacts (1-ProVar = 0.58), an adjustment for the second-order sampling error reveals the final, potentially true, correlation level $\hat{\rho}_{ir} = 0.163$. Our Hypothesis 2c is supported. Within CSP dimensions, CSP disclosure has the weakest relation to CFP, including second-order sampling error.

The Relation of CSP and Various CFP Categories

In line with Hypothesis 3a, perceptual performance yields – indeed, by some distance – the highest relation to CFP ($\hat{\rho}_i = 0.380$) (Table 4). It deviates significantly at the 99% level from the CFP group average ($\hat{\rho} = 0.148$), also adjusted for second-order sampling error ($\hat{\rho}_{ir} = 0.359$). This finding clearly supports Hypothesis 3a: CSP has the strongest link to perceptual CFP than to other CFP categories.

Operational CFP exhibits the highest CSP-CFP relation within the group of objective CFP categories ($\hat{\rho}_i = 0.222$) (Table 4). The deviation with the group average ($\hat{\rho} = 0.148$) is highly significant. Consideration of the second-order sampling error changes the result only negligibly ($\hat{\rho}_{ir} = 0.216$). This clearly supports Hypothesis 3b.

Accounting-based measures of CFP are more highly correlated ($\hat{\rho}_i = 0.161$) with CSP compared to other market-based measures: growth measures ($\hat{\rho}_i = 0.124$), market-based CFP measures ($\hat{\rho}_i = 0.115$), and risk measures ($\hat{\rho}_i = 0.062$).⁴ Even though all these category results are significantly different from zero, market-based CFP and risk measures return significantly lower correlations compared to other CFP measures (Table 4). These findings support Hypothesis 3c.

In accordance with Hypothesis 3d, we obtain the lowest correlation of all CFP categories ($\hat{\rho}_i = 0.016$) for CSP mutual funds. It is the only result within the CFP categories which is not significantly different from zero. Nonetheless, if we adjust for the low statistical reliability of the CSP mutual fund relation (1-ProVar = 0.00), a correction for the second-order sampling error towards the average group effect $\hat{\rho}$ becomes necessary (outlined in Equation 1). The difference between $\hat{\rho}_i$ and $\hat{\rho}_{ir}$ of 0.132 could be a reflection of the overall ‘noise’ – whether statistically or capital markets induced. We, thus, support Hypothesis 3d and conclude that mutual funds demonstrate a weaker CSP-CFP relation compared to other CFP categories. Figure 2 graphically summarizes the range and magnitude of the main effect sizes.

Moderating Factors

Our initial assessment reveals that meta-analyses which are more recent have significantly lower effect sizes compared to older studies. Measured by either the meta-study publication year ($\hat{\rho}_i = 0.130$ vs $\hat{\rho}_i = 0.216$) or the average

⁴For the purpose of this second-order meta-analysis, we changed the algebraic sign for CFP risk for calculation reasons. The stated positive CSP/CFP risk correlation means in fact a reduced CFP risk with increased CSP.

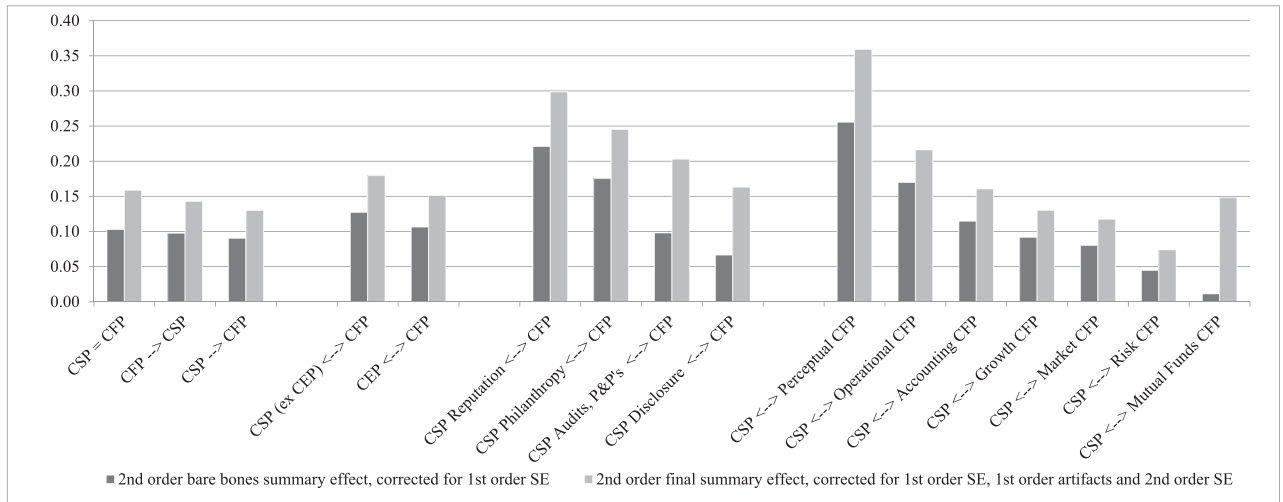


Figure 2. Graphical summary of second-order meta-analytical results for subgroups (effect sizes \bar{r}_i and \hat{p}_{ir}).

Hypothesis 4a

Average meta-analysis publication year

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	2013.2	62	1,252,187	0.092	0.003	0.130	0.005	0.004	0.008	0.113	0.146	0.000	0.259	4.84***	6.80***
H2	2005.6	67	1,135,028	0.140	0.003	0.216	0.006	0.006	0.009	0.197	0.234	0.064	0.368		

Average publication year of primary study in meta-analysis

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	2003.5	69	1,853,390	0.090	0.003	0.132	0.006	0.005	0.009	0.115	0.150	-0.013	0.277	4.33***	4.18***
H2	1991.4	55	437,839	0.137	0.004	0.194	0.008	0.008	0.012	0.171	0.218	0.021	0.368		

Average publication year of study (primary data sample, first-order meta-analysis)

sub	sub ρ	$n_{(1)}$	$N_{(1)}$	$\bar{r}_{(1)}$	$\sigma^2 \bar{r}_{(1)}$	$\hat{p}_{(1)}$	$S^2 \hat{p}_{(1)}$	$\hat{\sigma}^2 \hat{p}_{(1)}$	$SE \hat{p}_{(1)}$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_{(1)}$	$z \hat{p}_{(1)}$
H1	2006.1	291	90,324	0.120	0.026	0.168	0.056	0.050	0.013	0.143	0.195	-0.268	0.606	0.14	0.11
H2	1993.8	260	62,113	0.118	0.030	0.170	0.067	0.059	0.015	0.141	0.200	-0.307	0.648		

Hypothesis 4b

Number of primary studies in meta-analysis

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	151.5	64	1,612,851	0.103	0.002	0.153	0.006	0.005	0.009	0.135	0.170	0.012	0.294	1.09	0.54
H2	51.1	65	774,364	0.115	0.005	0.160	0.008	0.007	0.011	0.139	0.181	-0.008	0.329		

Hypothesis 4c

Orlitzky (2011) categorization

sub	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
GM journals	79	1,292,111	0.125	0.005	0.174	0.008	0.008	0.010	0.154	0.194	-0.001	0.349	3.8***	3.92***
SIM journals	22	288,082	0.082	0.001	0.115	0.003	0.003	0.011	0.093	0.137	0.011	0.218		

JCR classification

sub	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
Business and management	71	1,115,118	0.106	0.005	0.150	0.007	0.007	0.010	0.130	0.169	-0.012	0.311		
Envir. studies and ethics	11	262,766	0.081	0.001	0.113	0.003	0.003	0.015	0.083	0.143	0.013	0.212	1.87*	2.03**
Political science	4	95,490	0.160	0.002	0.224	0.004	0.004	0.031	0.162	0.285	0.101	0.347	2.28**	2.26**
Psychology, applied	7	55,046	0.155	0.000	0.197	0.001	0.001	0.010	0.177	0.216	0.145	0.248	4.24***	3.39***

Publishing journal impact factor

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	5.1	41	441,524	0.128	0.005	0.177	0.007	0.006	0.012	0.153	0.201	0.023	0.331	2.96***	3.27***
H2	1.8	52	1,086,896	0.091	0.002	0.127	0.005	0.005	0.009	0.108	0.145	-0.006	0.260		

Publishing status

sub	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
unpublished	28	807,022	0.115	0.002	0.186	0.004	0.004	0.012	0.163	0.209	0.064	0.308	0.63	2.42**
published	101	1,580,193	0.108	0.004	0.152	0.007	0.007	0.008	0.136	0.168	-0.011	0.314		

Paper citations per annum

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	196.7	26	317,247	0.115	0.002	0.159	0.004	0.004	0.013	0.134	0.184	0.034	0.284	0.59	0.14
H2	16.3	103	2,069,968	0.109	0.004	0.157	0.007	0.007	0.008	0.141	0.173	-0.009	0.323		

Hypothesis 4d

Overall methodological strength

sub	sub ρ	n_m	N_m	\bar{r}_i	$\sigma^2 \bar{r}_i$	\hat{p}_i	$S^2 \hat{p}_i$	$\hat{\sigma}^2 \hat{p}_i$	$SE \hat{p}_i$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrI_L^{95\%}$	$CrI_U^{95\%}$	$z \bar{r}_i$	$z \hat{p}_i$
H1	7.2	66	1,321,842	0.109	0.004	0.161	0.008	0.007	0.011	0.140	0.181	-0.008	0.329	0.07	0.51
H2	4.8	63	1,065,373	0.110	0.003	0.153	0.006	0.006	0.009	0.135	0.172	0.008	0.299		

Table 5. Moderating factors analysis.

Note: Unless otherwise stated, the subgroup H1 represents results greater than the respective average of the selected moderating factor. H2 represents the subgroup below the subgroup average. $z \bar{r}_i$ and $z \hat{p}_i$ are the corresponding z-values for the subgroup means. The additional index (1) indicates the first-order level. See text for abbreviations. The significance values are:

- * $p < 0.10$;
- ** $p < 0.05$; and
- *** $p < 0.01$.

publication year of the primary studies in each meta-analysis ($\hat{p}_i = 0.132$ vs $\hat{p}_i = 0.194$), the results indicate that meta-analyses which are more recent have significantly lower effect sizes (Table 5). However, the attenuation factors applied by the first-order meta-analytical reviewers for these two groups differ considerably. Relatively older meta-analyses used on average an attenuation factor of 0.66 in comparison to younger analyses (0.78). We also assess the relation on first-order level with data on 551 primary studies to rule out a distortion of initial results from changing meta-analytical methods and correction factors over time. This verification uncovers that the pattern of a correlation-level shrink is no longer observable as the differences in $\bar{r}_{(1)}$ and $\hat{p}_{(1)}$ do not exist anymore (0.002 and -0.002). Based on this extensive set of primary studies that avoids potential distortions through different meta-analytical attenuation factors, Hypothesis 4a cannot be confirmed.

When separating our meta-analytical results into dichotomous groups depending on the number of underlying studies (n), we retrieve two summary effects that are marginally and insignificantly different from each other. Larger studies tend to be negligibly weaker in effect size ($\hat{p}_i = 0.153$) in comparison to smaller studies ($\hat{p}_i = 0.160$). We conclude that the CSP-CFP relation is not dependent on the study sample size as the difference is non-significant. Larger studies in our sample produce equally strong relations compared to smaller samples which can be interpreted as another sign of the stability of the CSP-CFP relation. Thus, we reject Hypothesis 4b.

As no meta-study in our sample was published in an economics, finance, or accounting journal, we are limited to sub-effects in SIM journals and GM journals. We determine, in contrast to findings at primary study level, even significant lower effect sizes of meta-analyses in SIM journals ($\hat{p}_i = 0.115$) compared to GM journals ($\hat{p}_i = 0.174$). Concomitantly, we also find for the JCR journal classification significantly lower effect sizes in environmental studies and ethics compared to those in business and management ($\hat{p}_i = 0.113$ vs 0.150). These are interesting findings since prior research proposed the opposite. Working paper studies exhibit attenuated effect sizes comparable with those in published papers ($\bar{r}_i = 0.115$) but disattenuated effect sizes deviate significantly from those in published papers ($\hat{p}_i = 0.186$ vs 0.152). Moreover, we find that higher-impact factor journals (factor 5.1 on average for the sample) produce significantly higher correlations than lower-impact factor journals ($\hat{p}_i = 0.177$ vs $\hat{p}_i = 0.127$). Very

n_m	k_m	N_m	\bar{Y}	S_F^2	$\hat{\sigma}_F^2$	SE_F	$\hat{\beta}$	$S_{\hat{\beta}}^2$	$\hat{\sigma}_{\hat{\beta}}^2$	$SE_{\hat{\beta}}$	$CI_L^{95\%}$	$CI_U^{95\%}$	$CrL_L^{95\%}$	$CrL_U^{95\%}$	1-ProVar	Q	I^2	Fail-Safe N	z	\bar{z}	$z_{\hat{\beta}}$
(a)	25	4,507	992,239	0.108	0.001	0.001	0.150	0.002	0.002	0.009	0.132	0.169	0.058	0.242	0.96	11.4	0.00	987			
(b)	129	8,337	2,387,215	0.110	0.004	0.004	0.157	0.007	0.007	0.007	0.143	0.171	-0.001	0.316	0.96	75.5	0.00	1,861			
(c)	25	4,507	992,239	0.113	0.000	0.000	0.156	0.006	0.006	0.016	0.125	0.186	0.003	0.308	0.96	13.0	0.00	1,083	0.54	0.32	
(d)	129	8,337	2,387,215	0.117	0.004	0.004	0.193	0.014	0.013	0.010	0.173	0.213	-0.033	0.419	0.98	138.9	0.06	5,956	0.94	2.88***	
(e)	17	3,020	574,969	0.103	0.002	0.001	0.144	0.002	0.002	0.011	0.122	0.166	0.053	0.235	0.97	8.4	0.00	712	0.43	0.43	
(f)	54	4,273	1,257,796	0.113	0.004	0.004	0.163	0.006	0.006	0.011	0.141	0.184	0.007	0.318	0.97	37.5	0.00	1,338	0.29	0.41	
(g)	122	7,763	2,207,033	0.122	0.005	0.005	0.177	0.008	0.008	0.008	0.161	0.192	0.004	0.349	0.95	64.9	0.00	1,494	1.49	1.80*	
(h)	122	7,464	2,307,223	0.109	0.004	0.004	0.156	0.007	0.006	0.007	0.142	0.170	0.001	0.311	0.96	73.4	0.00	1,760	0.09	0.10	
(i)	122	7,995	2,191,187	0.112	0.004	0.004	0.160	0.007	0.006	0.007	0.146	0.174	0.005	0.315	0.96	70.0	0.00	1,973	0.33	0.29	
(j)	122	7,642	2,318,931	0.109	0.004	0.004	0.156	0.006	0.006	0.007	0.142	0.170	0.003	0.309	0.96	64.9	0.00	1,688	0.14	0.13	
(k)	118	7,678	2,291,690	0.105	0.003	0.003	0.151	0.006	0.005	0.007	0.138	0.164	0.008	0.294	0.95	55.8	0.00	1,546	0.60	0.63	

Table 6. Robustness analysis for summary effects ($m = 25$) and sub-effects ($m = 129$). Note: Table 6 displays the results of various robustness checks. The original results based on best-effort variances are (a) for $m = 25$ meta-analytic summary effects and (b) for $m = 129$ sub-effects. (c) Re-calculates (a) with simple variances instead of best-effort variances and (d) recalculates (b) with simple variances instead of best-effort variances. In particular, the comparison of (b) and (d) indicates that applied best-effort variances lead to more conservative estimates for effect sizes and narrower credibility intervals. The $\hat{\beta}$ -value for (d) is significantly higher compared to the corresponding value in (b) measured by its z-value. (e) and (f) display the results of additional measures to reduce overlapping in the sample, eliminating for (e) eight out of 25 meta-analyses in which more than 25% of primary studies overlap with each other, and also for (f) all sub-effects of those meta-analyses. We find no indication that small violations of independence affect effect sizes and credibility intervals despite the reduced data sample. From the $m = 129$ sub-effect sample, (g) eliminates the 5% of sub-effects with the lowest variances which would lead to significantly higher disattenuated effect sizes, (h) eliminates the 5% of sub-effects with the highest variances, (i) eliminates the 5% of sub-effects with the lowest effect sizes, (j) eliminates the 5% of sub-effects with the highest effect sizes, and finally (k) eliminates all sub-effects on reputational CSP and perceptual CFP. Viewed from different angles, the robustness checks underline once more the very robust nature of the CSP-CFP relation and the conservative nature of the determined effect sizes. See text for abbreviations. The significance values are:

* $p < 0.10$;
 ** $p < 0.05$; and
 *** $p < 0.01$.

frequently cited (per annum basis) meta-analyses exhibit very similar effect sizes compared to less often-cited ones. All these findings clearly reject Hypothesis 4c. There is no upwardly distorted CSP-CFP relation through SIM or lower-impact journals.

Finally, regarding the role of judgmental decisions in meta-analyses techniques and methodological choices, we find virtually no difference in effect sizes for methodological stronger ($\widehat{\rho}_i = 0.161$) and weaker studies ($\widehat{\rho}_i = 0.153$). This finding supports Hypothesis 4d. The CSP-CFP relation exists regardless of the methodological sophistication in meta-analyses. Furthermore, the results of the additional robustness tests suggest that the results are not distorted by statistical outliers (Table 6).

Discussion and Conclusion

This study makes several contributions to the academic inquiry on whether it pays to be a good firm. First, we are able to claim with more confidence than before that a highly significant, positive, and bidirectional CSP-CFP relation exists. The findings are robust at first-order and, in particular, at second-order meta-analytical level and do not change when we consider potential publication bias, apply less conservative variances, or conduct sensitivity analyses. True variances at second-order level are considerably lower compared to those at first-order level. Thus, the resulting credibility interval is much narrower, with a lower bound virtually at zero ($m = 129$). We are therefore safely able to claim that in 95% of cases we observe a positive average effect in meta-studies on the CSP-CFP relation. We find no indication that either CSP or CFP matters more from a cause–effect point of view. All three underlying theories (slack resources, good management, and virtuous circle) can empirically be defended. However, the level of effect sizes for a bidirectional relation is highest at the lowest variability of the effects. This leads us to acknowledge the virtuous circle theory as the most probable. To date, the academic CSP-CFP debate is often coined by the notions of ‘contradictory’, ‘mixed’, or ‘ambiguous’ findings. We propose that the business case for CSP can be considered well proved. Research designs taking into account non-linear CSP-CFP relations may even further increase the explanatory power of the CSP-CFP relation (Barnett & Salomon, 2012; Gao, Wu, & Hafsi, 2017; Trumpp & Guenther, 2017; Wagner, Van Phu, Azomahou, & Wehrmeyer, 2002).

Second, we shed light on an academically well-established misinterpretation. We cannot detect statistically significant differences between the effects of environmental and social-related CSP on CFP. This outcome is in contrast to several empirical studies and conceptions which consider social aspects as the main driver for CFP. Our findings propose a different result: good CSP pays off, whether social or environmental related. The difference between both is not statistically significant. This result mirrors S. L. Hart’s (1995) idea of a natural resource-based view and supports other empiric evidence for the business case of improving CEP (Albertini, 2013; Berchicci & King, 2007; Etzion, 2007). It may also be a way for revitalizing the resource-based view (Barney, Ketchen, & Wright, 2011) to acknowledge environmental- and social-related aspects as equally relevant sources for companies’ success.

Third, within CSP dimensions, it is possible to identify different sources for a business case. The strongest relation to CFP is identified for CSP reputation. The intangible character of corporate reputation, which is difficult to replicate and potentially the most overarching outcome measure for CSP, leads to sustained superior financial profit (Melo & Garrido-Morgado, 2012; P. W. Roberts & Dowling, 2002; Sabate & Puente, 2003). On the other hand, the strong correlation for CSP philanthropy reiterates that instrumental and moral initiatives can supplement each other (Brammer & Millington, 2008; Hahn, Pinkse, Preuss, & Figge, 2016). CSP disclosure as well as CSP audits, processes, and policies are significantly weaker correlations compared to the rest of the CSP dimensions. The limited standardization of CSP disclosure practices (Busch, Bauer, & Orlitzky, 2016) seems, in fact, to encourage companies to provide primarily beneficial rather than unbiased CSP information (Schultze & Trommer, 2012; Ullmann, 1985). The significantly weaker correlation of CSP disclosure may point to reliability and validity issues of CSP data. One way of overcoming such data limitations can be to strengthen efforts toward a more stringent – and maybe even obligatory – extra-financial data disclosure. Standardization efforts of the Sustainable Accountings Standards Board (SASB, 2015) and the European Union (EU, 2014) are promising steps forward in this regard.

Fourth, our findings suggest that operational performance is significantly and more highly correlated to CSP than other CFP categories. This finding mirrors other researchers' claims that operational performance such as productivity is located closer to the value creation source within a firm; it better reflects organizational performance than potentially distorted CFP categories like accounting- or market-based measures (McWilliams *et al.*, 1999; Pelozo, 2009; Rowley & Berman, 2000; Venkatraman & Ramanujam, 1986).

Fifth, it is an interesting phenomenon that correlations shrink for different CFP categories – perceptual performance having the highest correlation and CSP mutual funds the lowest. The closer research gets to realized market performance, the smaller the effects. Several reasons may explain why empirical studies on CSP mutual funds exhibit a relatively weak CSP-CFP relation. Realized performance in mutual funds depends on the overlapping effects of systematic and idiosyncratic risks (Campbell, Lettau, Malkiel, & Xu, 2001; Luo & Bhattacharya, 2009), construction constraints (Clarke, de Silva, & Thorley, 2002), and costs for portfolio implementation which can be as high as 2.5% per annum in various fees that are carried by the average mutual fund (Barber, Odean, & Zheng, 2005; Carhart, 1997; Khorana, Servaes, & Tufano, 2009). The 'drowned out by other noise' argument (Pelozo, 2009, p. 1527) thus seems to be even more valid for mutual funds, which are the ultimate amalgam of market-based measures. Nevertheless, in the worst case, mutual funds investors could expect to lose nothing compared to conventional mutual fund investments.

Sixth, we expected that early published analyses would show greater effect sizes compared with subsequently published meta-analyses. This potential diminishing effect over time as data accumulate is well documented (Trikalinos & Ioannidis, 2005). Indeed, we find significantly lower correlations for more recent meta-studies compared to older studies. However, the additional analysis at primary study level ($n = 551$) revealed that the shrinkage in correlations in the meta-analyses most likely does not indicate a shrinkage of the actual CSP-CFP relation. Correlations at primary study level, in contrast to our findings at the meta-study level, do not shrink. One explanation for this finding is that more recent meta-analytical research methods have become more sophisticated though using higher attenuation factors.

Finally, there was the suspicion raised that evidence of the CSP-CFP relation depends on the kind of publishing journal. The speculation was that SIM journals tend to publish 'better results'. Interestingly, we find the opposite result: 'truth is not made' in special SIM journals and low-impact factor journals, but rather, in particular, in high-impact factor and general management journals, at least at a meta-analysis level. We did not discover any publication bias analyzed by Fail Safe, funnel plot, or with regards to the publishing status of a paper. In line with previous findings (Aguinis *et al.*, 2011; Lipsey & Wilson, 1993), we also reveal that the methodological quality of meta-analyses has no influence on the reported effect sizes.

Concluding, this study takes an instrumental point of view: we ask whether it pays to be good and find that in fact it does. In light of the urgent global challenges that society faces, it appears essential that firms and investors consider any action possible to safeguard the living conditions for future generations. They can be encouraged that on first glance competing instrumental and moral motivations can supplement each other.

References

- Aguinis H, Glavas A. 2012. What We Know and Don't Know About Corporate Social Responsibility: A Review and Research Agenda. *Journal of Management* 38(4): 932–968.
- Aguinis H, Beaty JC, Boik RJ, Pierce CA. 2005. Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. *Journal of Applied Psychology* 90(1): 94–107.
- Aguinis H, Dalton DR, Bosco FA, Pierce CA, Dalton CM. 2011. Meta-Analytic Choices and Judgment Calls: Implications for Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact. *Journal of Management* 37(1): 5–38.
- Aguinis H, Gottfredson RK, Joo H. 2013. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods* 16(2): 270–301.
- Ahn S, Ames AJ, Myers ND. 2012. A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research* 82(4): 436–476.
- Albertini E. 2013. Does Environmental Management Improve Financial Performance? A Meta-Analytical Review. *Organization & Environment* 26(4): 431–457.
- Aldag RJ. 2013. No CPR for CSR: A Call to Abandon Search for the 'Holy Grail'. *Industrial and Organizational Psychology* 6(4): 372–376.
- Aldag RJ, Bartol KM. 1978. Empirical studies of corporate performance and policy: A survey of problems and results. In *Research in Corporate Social Performance and Policy* (Vol. 1), Preston LE (ed). JAI Press: Greenwich, CT; 165–199.

- Allouche J, Laroche P. 2005. A Meta-Analytical Investigation of the Relationship Between Corporate Social and Financial Performance. *Revue de Gestion des Ressources Humaines* 57: 18–41.
- AoM SIM Division. 2012. White Paper on A-Journal Lists. http://sim.aomonline.org/docs/SIM_WP_Journals.pdf
- Arlow P, Gannon MJ. 1982. Social Responsiveness, Corporate Structure, and Economic Performance. *Academy of Management Review* 7(2): 235–241.
- Aytug ZG, Rothstein HR, Zhou W, Kern MC. 2012. Revealed or Concealed? Transparency of Procedures, Decisions, and Judgment Calls in Meta-Analyses. *Organizational Research Methods* 15(1): 103–133.
- Bansal P, Clelland I. 2004. Talking trash: Legitimacy, impression management, and unsystematic risk in the context of the natural environment. *Academy of Management Journal* 47(1): 93–103.
- Barber BM, Odean T, Zheng L. 2005. Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows. *The Journal of Business* 78(6): 2095–2120.
- Barnett ML. 2007. Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *Academy of Management Review* 32(3): 794–816.
- Barnett ML, Salomon RM. 2012. Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal* 33(11): 1304–1320.
- Barnett ML, Jermier JM, Lafferty BA. 2006. Corporate Reputation: The Definitional Landscape. *Corporate Reputation Review* 9(1): 26–38.
- Barney JB. 1991. Firm Resources and Sustained Competitive Advantage. *Journal of Management* 17(1): 99–120.
- Barney JB, Ketchen DJ, Wright M. 2011. The Future of Resource-Based Theory: Revitalization or Decline? *Journal of Management* 37(5): 1299–1315.
- Baron DP. 2001. Private Politics, Corporate Social Responsibility, and Integrated Strategy. *Journal of Economics & Management Strategy* 10(1): 7–45.
- Bauer R, Koedijk K, Otten R. 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance* 29(7): 1751–1767.
- Berchicci L, King A. 2007. Postcards from the Edge: A review of the Business and Environment Literature. *The Academy of Management Annals* 1(1): 513–547.
- Bijmolt THA, Pieters RGM. 2001. Meta-Analysis in Marketing when Studies Contain Multiple Measurements. *Marketing Letters* 12(2): 157–169.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd.
- Borgers A, Derwall J, Koedijk K, ter Horst J. 2013. Stakeholder relations and stock returns: On errors in investors' expectations and learning. *Journal of Empirical Finance* 22(June): 159–175.
- Brammer S, Millington A. 2008. Does it pay to be different? An analysis of the relationship between corporate social and financial performance. *Strategic Management Journal* 29(12): 1325–1343.
- Buehler VM, Shetty YK. 1974. Motivations for Corporate Social Action. *Academy of Management Journal* 17(4): 767–771.
- Busch T, Bauer R, Orlitzky M. 2016. Sustainable Development and Financial Markets: Old Paths and New Avenues. *Business & Society* 55(3): 303–329.
- Campbell JY, Lettau M, Malkiel BG, Xu Y. 2001. Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance* 56(1): 1–43.
- Capelle-Blancard G, Monjon S. 2014. The Performance of Socially Responsible Funds: Does the Screening Process Matter? *European Financial Management* 20(3): 494–520.
- Carhart MM. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52(1): 57–82.
- Carroll AB, Shabana KM. 2010. The business case for corporate social responsibility: A review of concepts, research and practice. *International Journal of Management Reviews* 12(1): 85–105.
- Clarke R, de Silva H, Thorley S. 2002. Portfolio Constraints and the Fundamental Law of Active Management. *Financial Analysts Journal* 58(5): 48–66.
- Clarkson MBE. 1995. A Stakeholder Framework for Analyzing and Evaluating Corporate Social Performance. *Academy of Management Review* 20(1): 92–117.
- Cochran PL, Wood RA. 1984. Corporate Social Responsibility and Financial Performance. *Academy of Management Journal* 27(1): 42–56.
- Combs J, Liu Y, Hall A, Ketchen D. 2006. How much do high-performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology* 59(3): 501–528.
- Cooper H, Koenka AC. 2012. The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist* 67(6): 446–462.
- Crook TR, Todd SY, Combs JG, Woehr DJ, Ketchen DJ. 2011. Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology* 96(3): 443–456.
- Cyert RM, March JG. 1963. The behavioral theory of the firm. Prentice-Hall: Englewood Cliffs, NJ.
- Dalton DR, Daily CM, Certo ST, Roengpitya R. 2003. Meta-analyses of financial performance and equity: Fusion or confusion? *Academy of Management Journal* 46(1): 13–26.
- Daniel F, Lohrke FT, Fornaciari CJ, Turner RA. 2004. Slack resources and firm performance: a meta-analysis. *Journal of Business Research* 57(6): 565–574.
- Darnall N, Sides S. 2008. Assessing the Performance of Voluntary Environmental Programs: Does Certification Matter? *The Policy Studies Journal* 36(1): 95–117.
- Derwall J, Koedijk K, Ter Horst J. 2011. A tale of values-driven and profit-seeking social investors. *Journal of Banking & Finance* 35(8): 2137–2147.
- Devinney TM. 2009. Is the Socially Responsible Corporation a Myth? The Good, the Bad, and the Ugly of Corporate Social Responsibility. *Academy of Management Perspectives* 23(2): 44–56.

- Dickersin K. 2005. Publication Bias: Recognizing the Problem, Understanding Its Origins and Scope, and Preventing Harm. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, Rothstein HR, Sutton AJ, Borenstein M (eds). John Wiley & Sons, Ltd: Chichester, UK; 9–33.
- DiMaggio PJ, Powell WW. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48(2): 147–160.
- Dixon-Fowler HR, Slater DJ, Johnson JL, Ellstrand AE, Romi AM. 2013. Beyond ‘Does it Pay to be Green?’ A Meta-Analysis of Moderators of the CEP–CFP Relationship. *Journal of Business Ethics* 112(2): 353–366.
- Donaldson T, Preston LE. 1995. Theory The Stakeholder of the Concepts, Evidence, Corporation: and Implications. *Management* 20(1): 65–91.
- Eccles RG, Ioannou I, Serafeim G. 2014. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science* 60(11): 2835–2857.
- Egger M, Smith GD, Schneider M, Minder C. 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315(7109): 629–634.
- Elsayed K, Paton D. 2009. The impact of financial performance on environmental policy: does firm life cycle matter? *Business Strategy and the Environment* 18(6): 397–413.
- Endrikat J, Guenther E, Hoppe H. 2014. Making sense of conflicting empirical findings: A meta-analytic review of the relationship between corporate environmental and financial performance. *European Management Journal* 32(5): 735–751.
- Etzioni D. 2007. Research on Organizations and the Natural Environment, 1992–Present: A Review. *Journal of Management* 33(4): 637–664.
- European Union (EU). 2014. Directive 2014/95/EU of the European Parliament and of the Council. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0095>
- Fama EF. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25(2): 383–417.
- Fama EF. 1991. Efficient Capital Markets: II. *The Journal of Finance* 46(5): 1575–1617.
- Flammer C. 2015. Does Corporate Social Responsibility Lead to Superior Financial Performance? A Regression Discontinuity Approach. *Management Science* 61(11): 2549–2568.
- Fleming P, Jones MT. 2013. *The End of Corporate Social Responsibility: Crisis and Critique*. Sage Publications: London.
- Freeman RE. 1984. *Strategic management: A stakeholder approach*. Pitman: Boston.
- Friedman M. 1962. *Capitalism and Freedom*, 40th Anniv. The University of Chicago Press: Chicago and London.
- Friedman M. 1970. The social responsibility of business is to increase its profits. *The New York Times Magazine* September 13: 32–33.
- Frooman J. 1997. Socially Irresponsible and Illegal Behavior and Shareholder Wealth: A Meta-Analysis of Event Studies. *Business & Society* 36(3): 221–249.
- Fry LW, Keim GD, Meiners RE. 1982. Corporate Contributions: Altruistic or For-Profit? *Academy of Management Journal* 25(1): 94–106.
- Gao Y, Wu J, Hafsi T. 2017. The Inverted U-Shaped Relationship between Corporate Philanthropy and Spending on Research and Development: A Case of Complementarity and Competition Moderated by Firm Size and Visibility. *Corporate Social Responsibility and Environmental Management* 24(6): 465–477. <https://doi.org/10.1002/csr.1420>
- Glass GV. 1976. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* 5(10): 3–8.
- Godfrey PC, Merrill CB, Hansen JM. 2009. The relationship between corporate social responsibility and shareholder value: An empirical test of the risk management hypothesis. *Strategic Management Journal* 30(4): 425–445.
- Golicic SL, Smith CD. 2013. A meta-analysis of environmentally sustainable supply chain management practices and firm performance. *Journal of Supply Chain Management* 49(2): 78–95.
- Greenhouse JB, Iyengar S. 2009. Sensitivity analysis and diagnostics. In *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edition, Cooper H, Hedges LV, Valentine JC (eds). Russell Sage Foundation: New York; 417–433.
- Griffeth RW, Hom PW, Gartner S. 2000. A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium. *Journal of Management* 26(3): 463–488. <https://doi.org/10.1177/014920630002600305>
- Griffin JJ, Mahon JF. 1997. The Corporate Social Performance and Corporate Financial Performance Debate: Twenty-Five Years of Incomparable Research. *Business & Society* 36(1): 5–31.
- Guthrie JP. 2001. High-involvement work practices, turnover, and productivity: Evidence from New Zealand. *Academy of Management Journal* 44(1): 180–190.
- Hahn T, Pinkse J, Preuss L, Figge F. 2016. Ambidexterity for Corporate Social Performance. *Organization Studies* 37(2): 213–235.
- Hamilton S, Jo H, Statman M. 1993. Doing Well While Doing Good? The Investment Performance of Socially Responsible Mutual Funds. *Financial Analysts Journal* 49(6): 62–66.
- Hart SL. 1995. A Natural-Resource-Based View of the Firm. *Academy of Management Review* 20(4): 986–1014.
- Hart S, Ahuja G. 1996. Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment* 5: 30–37.
- Hattie J. 2009. *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge: New York.
- Hedges LV. 1981. Distribution Theory for Glass’s Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics* 6(2): 107–128.
- Hedges LV, Olkin I. 1985. *Statistical methods for meta-analysis*. Academic Press: San Diego.
- Higgins JPT. 2008. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* 37(5): 1158–1160.
- Higgins JPT, Thompson SG. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21(11): 1539–1558.
- Higgins JPT, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.)* 327(7414): 557–560.
- Hunter JE, Schmidt FL. 2004. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 2nd edition. Newbury Park, CA: Sage Publications.

- Hunter JE, Schmidt FL, Jackson GB. 1982. *Meta-analysis: cumulating research findings across studies*. Sage Publications: Beverly Hills, CA.
- Ioannidis JPA, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. 2003. Genetic associations in large versus small studies: an empirical assessment. *The Lancet* 361(9357): 567–571.
- Jiang RJ, Bansal P. 2003. Seeing the Need for ISO 14001. *Journal of Management Studies* 40(4): 1047–1067.
- Kazrin A, Durac J, Agteros T. 1979. Meta-meta analysis: A new method for evaluating therapy outcome. *Behaviour Research and Therapy* 17(4): 397–399.
- Khorana A, Servaes H, Tufano P. 2009. Mutual Funds Fees Around the World. *Review of Financial Studies* 22(3): 1279–1310.
- King DR, Dalton DR, Daily CM, Covin JG. 2004. Meta-analyses of post-acquisition performance: Indications of unidentified moderators. *Strategic Management Journal* 25(2): 187–200.
- Lee DD, Madhavan R. 2010. Divestiture and Firm Performance: A Meta-Analysis. *Journal of Management* 36(6): 1345–1371.
- Levitt T. 1958. The dangers of social responsibility. *Harvard Business Review* 36(5): 41–50.
- Light RJ, Pillemer DB. 1984. *Summing up. The science of reviewing research*. Harvard University Press: Cambridge, MA.
- Lipsey MW, Wilson DB. 1993. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist* 48(12): 1181–1209.
- Luo X, Bhattacharya C. 2009. The Debate over Doing Good: Corporate Social Performance, Strategic Marketing Levers, and Firm-Idiosyncratic Risk. *Journal of Marketing* 73(6): 198–213.
- Mackey JD, Frieder RE, Brees JR, Martinko MJ. 2017. Abusive supervision: A meta-analysis and empirical review. *Journal of Management* 43(6): 1940–1965.
- del Mar Miras-Rodríguez M, Carrasco-Gallego A, Escobar-Pérez B. 2015. Are Socially Responsible Behaviors Paid Off Equally? A Cross-cultural Analysis. *Corporate Social Responsibility and Environmental Management* 22(4): 237–256. <https://doi.org/10.1002/csr.1344>
- Margolis JD, Walsh JP. 2001. *People and profits?: The search for a link between a company's social and financial performance*. Psychology Press: New York.
- Margolis JD, Elfenbein HA, Walsh JP. 2009. Does it Pay to Be Good...And Does it Matter? A Meta-Analysis of the Relationship between Corporate Social and Financial Performance. *SSRN Electronic Journal* 1–68.
- Markowitz H. 1952. Portfolio Selection. *Journal of Finance* 7(1): 77–91.
- Markowitz H. 1959. *Portfolio Selection. Efficient Diversification of Investments*, 1970th edition. Yale University Press: New Haven and London.
- Mayer-Haug K, Read S, Brinckmann J, Dew N, Grichnik D. 2013. Entrepreneurial talent and venture performance: A meta-analytic investigation of SMEs. *Research Policy* 42(6–7): 1251–1273.
- McGuire JW. 1969. The Changing Nature of Business Responsibilities. *The Journal of Risk and Insurance* 36(1): 55–61.
- McGuire JB, Sundgren A, Schneeweis T. 1988. Corporate social responsibility and firm financial performance. *Academy of Management Journal* 31(4): 854–872.
- McWilliams A, Siegel D. 2000. Corporate Social Responsibility and Financial Performance: Correlation Misspecification? *Strategic Management Journal* 21(5): 603–609.
- McWilliams A, Siegel D. 2001. Corporate social responsibility: a theory of the firm perspective. *Academy of Management Review* 26(1): 117–127.
- McWilliams A, Siegel D, Teoh SH. 1999. Issues in the Use of the Event Study Methodology: A Critical Analysis of Corporate Social Responsibility Studies. *Organizational Research Methods* 2(4): 340–365.
- McWilliams A, Siegel DS, Wright PM. 2006. Corporate Social Responsibility: Strategic Implications. *Journal of Management Studies* 43(1): 1–18.
- Melo T, Garrido-Morgado A. 2012. Corporate Reputation: A Combination of Social Responsibility and Industry. *Corporate Social Responsibility and Environmental Management* 19(1): 11–31.
- O'Shaughnessy KC, Flanagan DJ. 1998. Determinants of layoff announcements following M&As: an empirical investigation. *Strategic Management Journal* 19(10): 989–999.
- Orlitzky M. 2001. Does Firm Size Confound the Relationship Between Corporate Social Performance and Firm Financial Performance? *Journal of Business Ethics* 33(2): 167–180.
- Orlitzky M. 2008. Corporate Social Performance and Financial Performance: A Research Synthesis. In *The Oxford Handbook of Corporate Social Responsibility*, 1st edition, Crane A, Matten D, McWilliams A, Moon J, Siegel DS (eds). Oxford University Press: New York; 113–134.
- Orlitzky M. 2011. Institutional Logics in the Study of Organizations: The Social Construction of the Relationship between Corporate Social and Financial Performance. *Business Ethics Quarterly* 21(3): 409–444.
- Orlitzky M. 2013. Corporate social responsibility, noise, and stock market volatility. *Academy of Management Perspectives* 27(3): 238–254.
- Orlitzky M, Benjamin JD. 2001. Corporate Social Performance and Firm Risk: A Meta-Analytic Review. *Business & Society* 40(4): 369–396.
- Orlitzky M, Schmidt FL, Rynes SL. 2003. Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies* 24(3): 403–441.
- Pavie JJE, Filho LANC. 2008. Corporate social responsibility and financial performance: a meta-analysis. 34th EIBA annual conference paper.
- Pelozo J. 2009. The Challenge of Measuring Financial Impacts From Investments in Corporate Social Performance. *Journal of Management* 35(6): 1518–1541.
- Peterson RA. 2001. On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research* 28(3): 450–461.
- Porter ME. 1979. How competitive forces shape strategy. *Harvard Business Review* 57(2): 137–145.
- Rathner S. 2013. The Influence of Primary Study Characteristics on the Performance Differential Between Socially Responsible and Conventional Investment Funds: A Meta-Analysis. *Journal of Business Ethics* 118(2): 349–363.
- Revelli C, Viviani J-L. 2015. Financial performance of socially responsible investing (SRI): what have we learned? A meta-analysis. *Business Ethics: A European Review* 24(2): 158–185.
- Richard D, Bond CF, Stokes-Zoota JJ. 2003. One Hundred Years of Social Psychology Quantitatively Described. *Review of General Psychology* 7(4): 331–363.

- Richard PJ, Devinney TM, Yip GS, Johnson G. 2009. Measuring Organizational Performance: Towards Methodological Best Practice. *Journal of Management* 35(3): 718–804.
- Roberts RW. 1992. Determinants of corporate social responsibility disclosure: An application of stakeholder theory. *Accounting, Organizations and Society* 17(6): 595–612.
- Roberts PW, Dowling GR. 2002. Corporate reputation and sustained superior financial performance. *Strategic Management Journal* 23(12): 1077–1093.
- Romano JL, Kromrey JD. 2008. What Are the Consequences If the Assumption of Independent Observations Is Violated in Reliability Generalization Meta-Analysis Studies? *Educational and Psychological Measurement* 69(3): 404–428.
- Rosenbusch N, Bausch A, Galander A. 2007. The impact of environmental characteristics on firm performance: A meta-analysis. *Academy of Management Annual Meeting Proceedings* 8(1): 1–6.
- Rosenbusch N, Brinckmann J, Bausch A. 2011. Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing* 26(4): 441–457.
- Rosenthal R. 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin* 86(3): 638–641.
- Rowley T, Berman S. 2000. A Brand New Brand of Corporate Social Performance. *Business & Society* 39(4): 397–418.
- Rubera G, Kirca AH. 2012. Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of Marketing* 76(3): 130–147.
- Sabate JMDLF, Puente EDQ. 2003. Empirical Analysis of the Relationship Between Corporate Reputation and Financial Performance: A Survey of the Literature. *Corporate Reputation Review* 6(2): 161–177.
- SASB. 2015. SASB Annual Report 2014. San Francisco.
- Schmidt FL, Hunter JE. 2015. *Methods of meta-analysis: correcting error and bias in research findings*, 3rd edition. Sage Publications: Newbury Park, CA.
- Schmidt FL, Oh I-S. 2013. Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Processes* 121(2): 204–218.
- Schultze W, Trommer R. 2012. The concept of environmental performance and its measurement in empirical studies. *Journal of Management Control* 22(4): 375–412.
- Stam W, Arzlanian S, Elfring T. 2014. Social capital of entrepreneurs and small firm performance: A meta-analysis of contextual and methodological moderators. *Journal of Business Venturing* 29(1): 152–173.
- Sutton AJ. 2009. Publication Bias. In *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edition, Cooper H, Hedges LV, Valentine JC (eds). Russell Sage Foundation: New York; 435–452.
- Tamim RM, Bernard RM, Borokhovski E, Abrami PC, Schmid RF. 2011. What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research* 81(1): 4–28.
- Tracz SM, Elmore PB, Pohlmann JT. 1992. Correlational Meta-Analysis: Independent and Nonindependent Cases. *Educational and Psychological Measurement* 52(4): 879–888.
- Trikalinos TA, Ioannidis JPA. 2005. Assessing the Evolution of Effect Sizes over Time. In *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, Rothstein HR, Sutton AJ, Borenstein M (eds). John Wiley & Sons, Ltd: Chichester, UK; 241–259.
- Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, Ioannidis JPA. 2004. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *Journal of Clinical Epidemiology* 57(11): 1124–1130.
- Trumpp C, Guenther T. 2017. Too Little or too much? Exploring U-shaped Relationships between Corporate Environmental Performance and Corporate Financial Performance. *Business Strategy and the Environment* 26(1): 49–68.
- Tubre TC, Collins JM. 2000. Jackson and Schuler (1985) Revisited: A Meta-Analysis of the Relationships Between Role Ambiguity, Role Conflict, and Job Performance. *Journal of Management* 26(1): 155–169.
- Ullmann AA. 1985. Data in Search of a Theory: A Critical Examination of the Relationships among Social Performance, Social Disclosure, and Economic Performance of U. S. Firms. *The Academy of Management Review* 10(3): 540–557.
- Unger JM, Rauch A, Frese M, Rosenbusch N. 2011. Human capital and entrepreneurial success: A meta-analytical review. *Journal of Business Venturing* 26(3): 341–358.
- Venkatraman N, Ramanujam V. 1986. Measurement of Business Performance in Strategy Research: A Comparison of Approaches. *The Academy of Management Review* 11(4): 801–814.
- Venkatraman N, Ramanujam V. 1987. Measurement of Business Economic Performance: An Examination of Method Convergence. *Journal of Management* 13(1): 109–122.
- Vishwanathan P. 2010. The elusive relationship between corporate social and financial performance: Meta-analyzing four decades of misguided evidence. *Academy of Management Proceedings* 2010(1): 1–7.
- Waddock SA, Graves SB. 1997. The Corporate Social Performance-Financial Performance Link. *Strategic Management Journal* 18(4): 303–319.
- Wagner M, Van Phu N, Azomahou T, Wehrmeyer W. 2002. The relationship between the environmental and economic performance of firms: an empirical analysis of the European paper industry. *Corporate Social Responsibility and Environmental Management* 9(3): 133–146.
- Wall TD, Michie J, Patterson M, Wood SJ, Sheehan M, Clegg CW, West M. 2004. On the validity of subjective measures of company performance. *Personnel Psychology* 57(1): 95–118.
- Wernerfelt B. 1984. A resource-based view of the firm. *Strategic Management Journal* 5(2): 171–180.
- van Wijk R, Jansen JJP, Lyles MA. 2008. Inter- and Intra-Organizational Knowledge Transfer: A Meta-Analytic Review and Assessment of its Antecedents and Consequences. *Journal of Management Studies* 45(4): 830–853.
- Wilson DB, Lipsey MW. 2001. The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods* 6(4): 413–429.

- Wiseman J. 1982. An evaluation of environmental disclosures made in corporate annual reports. *Accounting, Organizations and Society* 7(1): 53–63.
- Wood DJ. 2010. Measuring Corporate Social Performance: A Review. *International Journal of Management Reviews* 12(1): 50–84.
- Wu M-L. 2006. Corporate Social Performance, Corporate Financial Performance, and Firm Size: A Meta-Analysis. *The Journal of American Academy of Business* 8(1): 163–171.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.