ESG and financial performance: aggregated evidence from more than 2000 empirical studies

# Introduction

Thus, there is an ongoing debate about the role and the impact of the financial sector on the natural environment and society (Weber 2014). In order to derive a more comprehensive picture, several review studies summarize primary ESG–CFP studies. Yet, all these first-level review studies provide an incomplete picture. This study is the first effort to provide aggregated evidence based on more than 2000 empirical studies that have been released since the 1970s.

We chose a two-step research method to analyze existing review and primary studies. First, we include findings from so-called vote-count studies. Vote-count studies count the number of studies with significant positive, negative, and nonsignificant results and “votes” the category with the highest share as winner (Light and Smith 1971).

Second, we aggregate the findings of econometric review studies – so-called meta-analyses – to derive a second-order meta-analysis

Chart, histogram

Description automatically generated

# Data

# Search

# Sample

we identified 35 vote-count studies (Table 1) and 25 meta-analyses (Table 2) which combine results from 3718 (gross) primary studies of which 1816 studies stem from votecount studies and 1902 from meta-analyses.

# Method

Depending on data availability in vote-count studies and meta-analyses, an analysis for different asset classes, regions, categories of E, S, and G as well the relation over time is conducted. When both vote-count studies and meta-analyses offer this information, the more comprehensive primary study sample is chosen. When the sub-sample stems from vote-count studies, the analysis focuses on the distribution of outcomes; when the sub-sample stems from meta-analyses, the focus is on effect sizes. Raw correlations, corrected correlations, sample sizes as well as corresponding variances, standard errors, confidence interval (CI), and credibility interval (CrI) have been extracted from the original meta-analyses as far as possible for further calculations

## Calculation of distributions

Vote-count studies

Distributions of positive, negative, neutral, and mixed outcomes are calculated for vote-count studies based on the results of the gross study sample and the net study sample.

On average, every unique primary study in the vote-count sample is analyzed by 1.8 review authors. To decide on the overall interpretation per unique study, a binomial test with three equally probable outcomes is applied (positive, neutral, and negative). A probability of greater than .95 served as cut-off point to determine the final interpretation for the study. If no clear positive or negative assignment was possible, the study is classified as neutral and/or mixed

Meta Analysis

Vote-count reviewers provide an assessment of the extent to which an observed relation in a primary study is a significant outcome. When undertaking a meta-analysis of primary studies, this assessment is performed by the second-level reviewer.

The true variance for r =ˆ i (the meta-analytical mean of attenuated correlations) is sˆ 2 r =ˆ i and the corresponding variance for p =ˆ i (the meta-analytical mean of disattenuated correlations) is sˆ 2p =ˆ i . As we are interested in the degree of significant positive and negative results, we place the intervals around zero instead of the meta-analytical mean

## Calculation of effect sizes

Vote-count studies

Even though vote-count studies usually do not report effects sizes like standardized mean differences (d/g) or correlations, it is possible to approximate them with the provided data.

The effect size r is determined by calculating the ratio of p0(r), which divides the number of positive studies by the sum of positive and negative studies, and by putting it in relation with the corresponding number of studies n.

Meta Analysis

All other average effect sizes and summary statistics of the 25 meta-analyses are determined with Schmidt and Oh’s method for second-order meta-analysis (Schmidt and Oh 2013). A secondorder meta-analysis combines a number of methodologically comparable and independent firstorder meta-analyses. It allows knowledge aggregation across a tremendous set of primary studies

In order to differentiate whether correlations and corresponding variance are first-order (based on extracted primary studies) or second-order (aggregated vote-count studies or meta-analyses), we add one or two lines above letters for attenuated correlations r and disattenuated correlations p.

## Summary effects: distributions

Vote-count studies

In a first step for the analysis of distribution results, all 1816 vote-count studies in the gross sample are treated as unique studies without adjusting for overlap among the vote-count studies. The overall weighted share of positive findings in the sample is calculated at 48.2%. In 41.0% of all results, the findings lead to neutral (23.0%) or mixed findings (18.0%). Just 10.7% of all analyzed studies exhibit a negative ESG–CFP relation.

Meta Analysis

Out of the 25 meta-analyses in the sample, just one study displays a summary effect size that has a negative ESG–CFP correlation – albeit very close to zero.

## Summary Effects: Correlations

Vote-count studies

Next, an approximation of the correlation effect size in vote-count studies based on the vote-count method of Hedges and Olkin (1985, 47ff) is conducted. The weighted average correlation rv = in all vote-count studies is calculated at 0.146. The corresponding p-value of ndicates a correlation factor highly significant and different from zero.

Meta-analyses

For reasons of comparability with the vote-count effect size estimate rv =, we compute the attenuated sample-size weighted average correlation for the 25 meta-analyses. The calculated correlation rm = is 0.118. The p-value of similarity of rm = and rv = is notably high at 0.638. This means vote-count studies and meta-analyses determine statistically comparable results for the ESG– CFP relation. However, generalizing this finding for both methods universally may not be appropriate due to the almost independent samples containing few overlaps and very different variance levels.

Portfolio studies and nonportfolio studies

The relevance of this distinction becomes apparent when looking at the vote-count studies. The share of positive results in the n = 155 identifiable portfolio-related studies shrinks considerably (15.5%) in comparison to nonportfolio-based studies (56.7%). Studies with neutral or mixed findings increase proportionately in portfolio-based studies and constitute nearly three quarters. The share of negative studies increases marginally compared to nonportfolio studies (11.0% vs. 5.8%) (Figure 4). Comparable results are found when all portfolio-focused vote-count studies are separately analyzed based on estimated effect sizes. The five primarily portfolio-focused vote-count studies exhibit a negative correlation rv = (p) of –0.061 in comparison to the 30 primarily nonportfolio-focused vote-count studies with rv = (non-p) of 0.177. The difference between both groups is highly significant (Table 5).

Table

Description automatically generated

Sub-effects in asset classes

nonequity asset classes both for bonds and real estate display a considerably higher share of positive findings over equities. More than two-thirds of studies uncover significant positive performance relations to ESG criteria. The share of positive votes for the 36 analyzed bond studies stands at 63.9% – with 13 neutral or mixed findings (36.1%).

Chart, bar chart, waterfall chart

Description automatically generated

Sub-effects in ESG categories

A key question is whether any of the three ESG letters may have a dominating effect on CFP. Some meta-analyses find significant positive relations for corporate environmental performance and CFP. Human capital-focused meta-analyses also find highly significant positive correlations.

However, not all of the E-, S-, and G-specific findings are free from ambiguity and no largescale comparison between the subgroups has been undertaken yet.

When reviewing studies with various combinations of ESG criteria, 35.3% report positive (respectively 7.1% negative) findings. The downside bias primarily arises from a high proportion of portfolio-based studies in this section (39.1%). If all these studies were excluded, the positive (negative) rate stands at 51.7% (4.8%) which is nonetheless lower than pure E, S, and G approaches (Figure 6)

Chart, waterfall chart

Description automatically generated

Sub effects of regions

Chart, waterfall chart

Description automatically generated

ESG effect over time

The question has been raised of whether the ESG–CFP relation is stable over time (Griffin and Mahon 1997; Borgers et al. 2013). Theoretically, the increasing amount of UN PRI signatories and, presuming an increasing ESG awareness within investment strategies, a decreasing ESG alpha (shrinking correlations over time) would be expected due to learning effects in capital markets. Empiric findings of meta-analyses investigating if investors’ increased focus onstakeholder issues also lead to changing ESG–CFP patterns over time present a fuzzy picture

# Discussion

Both vote-count and meta-analytic studies yield comparable results. This is a surprising outcome since the underlying studies are comprised of nearly independent samples (12.9% overlap) and apply different methods. Both methods yield robust results which reinforces the claim that there is a business case for ESG investing.

While overall correlation averages between 0.108 (r =ˆ i) and 0.169 ( pˆ i) could be considered rather “small” (Cohen 1988, 1992), they reflect common effect sizes in social sciences (Richard, Bond, and Stokes-Zoota 2003; Tamim et al. 2011; Lipsey et al. 2012) and, notably, might have relatively high relevance for competitive global securities markets.

We argue that ESG portfolios should be expected to exhibit lower correlations to CFP and less positive findings for the following three reasons: (1) following the “drowned out by noise” argument (Peloza 2009), various overlapping market and nonmarket factors in a portfolio tend cover potentially existing ESG alpha. (2) Most ESG funds constitute a mixture of so-called negative and positive ESG-screened funds, which could result in distortion and cancellation of any remaining effects (Derwall, Koedijk, and Ter Horst 2011). (3) Only studies on portfolios (in particular mutual funds) embed management fees and other costs such as performance fees and trading costs.

. Investors, on average, are unlikely to harvest the existing ESG alpha after implementation costs. However, it can be argued, sophisticated investors are more likely to do so Thus, our results underpin previous findings: at the worst case, investors in ESG mutual funds can expect to lose nothing compared to conventional fund investments

# Conclusion

A key area for future research is to better understand the interaction of different ESG criteria in portfolios and the relevance of specific ESG sub-criteria for CFP. These insights will shed further light on the ESG determinants for long-term positive performance impacts