On the Relation Between CSR and Financial Performance

Amrou Awaysheh*
Randall A. Heron
Tod Perry
Jared I. Wilson

* corresponding author, awaysheh@iu.edu

Affiliation for all authors:

Kelley School of Business
Indiana University
Indianapolis, IN 46202

Citation details:

Awaysheh Amrou, Heron Randall A., Perry Tod, Wilson Jared I, (2020) *forthcoming*. On the Relation Between CSR and Financial Performance" Strategic Management Journal, https://doi.org/10.1002/smj.3122

This is the author's manuscript of the article published in final edited form as:

Awaysheh, A., Heron, R. A., Perry, T., & Wilson, J. I. (2020). On the relation between corporate social responsibility and financial performance. Strategic Management Journal, 41(6), 965–987. https://doi.org/10.1002/smj.3122

On the Relation Between CSR and Financial Performance

Abstract

Research Summary:
This study reexamines the relation between CSR and financial performance by benchmarking firms against industry peers in a given year to identify *best-in-class* and *worst-in-class* firms. We also address distributional issues when using CSR ratings (clustering of CSR scores around the median and material differences across industries and time) and financial performance ratios (the possible influence of extreme values). We find that the *best-in-class* firms outperform their industry peers in terms of operating performance and have higher relative market valuations (Tobin's Q). When we control for endogeneity, we find that the significant relation between operating performance and CSR categories disappears, calling into question whether this relation is causal. However, we continue to find that *best-in-class* firms receive higher relative market valuations than industry peers.

Managerial Summary: The conflicting evidence on the relation between CSR and firm performance may influence a manager's decision to invest in CSR activities and an investor's decision to invest in a firm. Our research provides managers and investors with important implications regarding the value of relative benchmarking. Managers should understand that expectations of CSR performance evolve over time and that investors place higher valuations on the *best-in-class* CSR firms within an industry.

Keywords: CSR; firm performance; MSCI KLD; operating income; firm value

Introduction

Corporate Social Responsibility (CSR) continues to be a hot topic for firms, the business press, and the investment community. Many of the largest firms around the world not only invest significant resources implementing CSR initiatives, but also work to disclose their CSR activities to various stakeholders and potential investors through different channels including annual sustainability reports (Wang, Hsieh, and Sarkis, 2018). Firms are also incorporating CSR criteria, such as adding the cost of carbon or the cost of water in financial projections, when assessing the merits of capital allocation decisions (Flammer, 2015).

The investment community has responded as well, with many funds now explicitly considering CSR among various investment criteria. The US Sustainable Investment Foundation reports that the number of investment funds incorporating Environmental, Social, and Governance (ESG) factors in their investment decisions has grown dramatically, with over $11.6 trillion of assets under management in 2018 (SIF, 2018). In addition, “Green Bonds” are becoming more prevalent to finance projects with a CSR focus among firms where environmental issues are particularly acute (Flammer, 2018), with over $120 billion in new Green Bond issuances in 2017 and $500 billion in 2018 (World Bank, 2017; 2018). Given the increased attention to CSR from financial markets, it is not surprising that firms with good CSR performance appear to have better access to capital (Cheng, Ioannou, and Serafeim, 2014).

Related literature on value of CSR

Despite evidence that CSR is increasingly considered to be important by firms and capital market participants, there remains an unsettled debate among researchers regarding the relation between CSR and firm performance. Friedman (1970) framed one side of the theoretical debate decades ago, suggesting that CSR investments that ultimately benefit other stakeholders at the expense of shareholders will lead to reduced corporate profits and stock prices. An alternative view, however, is that CSR investments can generate a sustainable competitive advantage that allows firms improved financial performance that is valued by financial markets and benefits shareholders. The empirical evidence is, at best, mixed. For example, in an early and influential study of the relation between CSR and financial performance, Waddock and Graves (1997) report a positive relation between CSR ratings and prior financial performance and a positive relation between CSR ratings and subsequent financial performance. McWilliams and Siegel (2000) suggest that Waddock and Graves' model suffered from an omitted variable problem since it did not include R&D intensity, which is highly correlated with CSR. They show that when R&D intensity is included, the positive relation between CSR and subsequent financial performance disappears. In addition, a recent replication study using a much larger sample across a longer timeframe conducted by Zhao and Murrell (2016) confirms the original finding that prior financial performance relates to subsequent corporate social performance, but "casts doubt" on whether there still exists a positive relation between CSR and subsequent financial performance.

Studies that examine the value of CSR by focusing on announcement returns associated with additions or deletions to indices comprised of the top CSR firms (Durand, Paugam, and Stolowy, 2019; Hawn, Chatterji, and Mitchell, 2018; Ramchander, Schwebach, and Staking, 2012) also report mixed results and conclusions. If investors are not aware of the significance of the CSR activities or are unable to distinguish the quality of such activities, then external monitors may provide valuable benefits to investors through a certification role. Ramchander *et al*. (2012) examine announcement returns associated with additions and deletions to the Domini Social 400 Index (DS400) and find positive wealth effects for firms added to the DS400 and negative contemporaneous wealth effects for rival firms, suggesting that the DS400 certifies the quality of stakeholder-centered CSR for these firms. Hawn *et al*. (2018), however, examine investor reactions to additions, deletions, or continuation on the DJSI World index and conclude that DJSI events have an immaterial effect on firm value after including control variables. Durand *et al*. (2019) confirm Hawn *et al*.'s (2018) findings regarding stock price reactions to DSJI events, but also find that DSJI sustainability events lead to greater analyst coverage and increases in shareholdings by long-term investors. Security analysts appear to play an important role in linking corporate social performance to financial performance by reducing the information asymmetry between firms and investors related to investments in CSR (Luo, Wang, Ratihel, and Zheng, 2015).

Nonetheless, many researchers have reported that firms and their investors benefit from CSR activities when examining the role of CSR in specific contexts. For example, investment in CSR can mitigate risk and build reputational capital (Godfrey, Merrill, and Hansen, 2009), function as an insurance mechanism (Godfrey, 2005; Koh, Qian, and Wang, 2014; Shiu and Yang, 2017), or benefit the firm through the "halo effect" when the firm is subject to sanctions (Hong and Liskovich, 2015). Disclosure of CSR investments indicating greater stakeholder engagement and transparency may also increase the access to capital for firms (Cheng *et al*., 2014), particularly in times when capital may be limited or in crisis periods when social capital and trust is more important (Amiraslani, Lins, Servaes, and Tamayo, 2018; Lins, Servaes, and Tamayo, 2017).

Measurement of social and financial performance

There are several possible explanations for the inconclusive findings regarding the relation between CSR and financial performance in the extant literature. One possibility that we subscribe to in this study, is that it is difficult for researchers to accurately measure CSR and conduct the appropriate tests (Chen and Delmas, 2011). First, finding appropriate data on CSR performance at the firm level is difficult, and ratings provided by different agencies do not always converge

(Chatterji, Durand, Levine, and Touboul, 2016). Firms do not fully disclose all actions, and the nature of ESG activities contributes to the difficulty of measuring and comparing across time, firms, and industries. Thus, researchers have relied on third party sources to measure a firm's CSR performance or to provide external certification, such as MSCI KLD, Asset4, Dow Jones Sustainability Index, or Sustainalytics. Second, these CSR data are typically organized and reported in a way that we believe requires a more nuanced analytical approach. For example, we describe and show how time trends and industry characteristics embedded in CSR ratings should be considered when analyzing the data. Moreover, we show that significant clustering exists around the median CSR scores for a given industry. We propose that because of the time trends, industry differences, clustering, and the potential for a nonlinear relation between CSR and financial performance (Barnett and Solomon, 2012), contrasting firms that are *best-in-class* with other firms, including *worst-in-class* is a better way to analyze and interpret the relation between CSR and financial performance. Rather than concluding a significant positive relation, or perhaps a curvilinear relation exists, researchers can communicate more succinctly that *best-in-class* firms outperform *worst-in-class* firms (or all other peers) by a certain amount.

Several studies of sustainability and CSR have previously drawn comparisons between firms along the lines of *best-in-class* versus *worst-in-class.* As one example, Eccles, Ioannou, and Serafeim (2014) use propensity score matched samples of *High* and *Low* sustainability firms and find that the high sustainability firms outperform the low sustainability firms in terms of both stock price and accounting performance. Although there are several possible ways to categorize high/low, best-in-class/worst-in-class, etc. (e.g., median cuts, comparisons of top x percent vs bottom x percent), 10 percent thresholds are commonly used to discern between best and worst in class. For example, the Dow Jones Sustainability Index (DJSI) limits inclusion to those firms in the top 10 percent of the industry, based on its sustainability scores. Kempf and Osthoff (2007) construct and compare portfolios consisting of the top 10 percent (high-rated portfolio) and bottom 10 percent (low-rated portfolio) of KLD social responsibility ratings and report that a strategy of buying a portfolio of high-rated stocks and shorting the portfolio of low-rated stocks leads to significantly positive abnormal returns during their sample period. As another example, Verheyden, Eccles, and Feiner (2016) use a 10 percent threshold for *excluding* firms that are worst-in-class and find that the returns for portfolios that exclude the bottom 10 percent of the ESG universe outperformed those of the universe portfolio without negatively affecting the benefits of diversification.[1] Because of the significant clustering around the median noted earlier and the previously discussed research, we use 10 percent thresholds to categorize *best-in-class* and *worst-in-class* in this study.

A third challenge to analyzing the relation between CSR and financial performance results from the extreme values that often arise when using variables calculated as ratios (such as ROA and Tobin's Q) to capture financial performance (Certo, Busenbark, Kalm, and Lepine, 2018). Unless researchers take appropriate measures to prevent outlier ratios from distorting means and regression results, a few extreme values may significantly influence analyses and conclusions. Because these measures of financial performance are central to our analysis and comparability to Waddock and Graves (1997) and Zhao and Murrell (2016), we discuss below and in further detail in the internet appendix the steps we take to prevent outliers from driving our results.

[1] Distinguishing between firms that are *best-in-class* or *worst-in-class* is also consistent with recent discussion in the business press that investors are more interested in "companies that have the best financial performance among those with positive ESG ratings, reflecting a change in emphasis among investors." *Do-Good ETFs Finally Pay Off for Investors.* By Dan Weil, Wall Street Journal, May 7, 2018.

The relation between CSR and financial performance

Finally, we contend that the relation between CSR and financial performance can be difficult to disentangle because of possible reverse causality, omitted variable, and endogeneity concerns. To demonstrate the importance of this issue, we conduct a series of analyses, beginning with univariate tests, followed by a multivariate analysis that incorporates control variables known to be related to our performance measures, and concluding with an instrumental variable two-stage least squares multivariate analysis to address causality and endogeneity. Our instrumental variables approach utilizes location-based instruments that take into account the concentration of nearby firms that are *best-in-class*. Our choice of location-based instruments is based on the literature establishing relevance through a linkage between local CSR density and CSR engagement (Husted, Jamali, and Saffar, 2016), and that has operationalized location-based instruments to mitigate endogeneity problems (Karpoff, Schonlau, and Wehrly, 2017).

Although the evidence regarding the connection between CSR and financial performance discussed earlier is mixed, we contend that the numerous scenarios identified by researchers in which CSR leads to improved financial outcomes should allow for the identification of a positive relation between CSR and firm performance and valuation. This leads us to hypothesize that *best-in-class* CSR firms within an industry will have better financial performance and higher overall valuations than their *worst-in-class* (and other) peers. Our contribution to the literature is to quantify how *best-in-class* firms compare to their *worst-in-class* (and other) industry peers while taking the appropriate steps to deal with time trends and the median clustering present in the CSR data, making necessary adjustments to prevent extreme values of financial performance ratios from distorting our results, and addressing endogeneity concerns.

Our univariate and initial multivariate analyses that do not account for endogeneity provide evidence supporting the position that firms can do well by doing good. Specifically, we show that *best-in-class* firms with the highest CSR performance (top 10 percent) relative to industry peers in a given year have higher levels of operating performance (measured by Operating Income Before Depreciation (OIBD) scaled by Total Assets) and higher relative valuations (measured by Tobin's Q). However, we caution that the positive relation between operating performance and CSR does not imply that spending more on CSR increases operating performance. Instead, it is plausible that firms that are more profitable are able to invest more on CSR activities without raising the ire of shareholders. It is also possible that the higher relative market valuations (Tobin's Q) for firms that are *best-in-class* result from their higher levels of operating performance.

When controlling for endogeneity with location-based instruments, we continue to find that *best-in-class* firms have higher relative market valuations (Tobin's Q); however, we no longer find that *best-in-class* firms outperform their industry peers in terms of operating performance. This illustrates that controlling for endogeneity is important when interpreting the causal relation between CSR and firm performance. In particular, as McWilliams and Siegel (2000) suggested when revisiting the Waddock and Graves (1997) study, omitted variables that correlate with financial performance measures can lead to biased inferences. While our multivariate analysis reported prior to the instrumental variable approach includes a host of variables commonly used as controls when studying financial performance, it is apparent that important omitted variables remain, necessitating the need to take further steps to address the endogeneity issues. In addition to the complications that can arise when analyzing CSR data and studying financial ratios, endogeneity issues quite possibly explain why the literature that investigates the relation between CSR and financial performance has produced such mixed results.

We organize the remainder of this study as follows. In the next section, we describe the data and detail the methodology used in our study. In the subsequent section, we report results from the univariate, multivariate, and instrumental variable multivariate analyses, and then conclude by discussing our findings and possible limitations.

Data and Methodology

Methods

We utilize the Fama-French 48 industry classifications (FFI48) to facilitate comparisons of firms operating in the same industry (Fama and French, 1997). As measures of financial performance, we use both accounting operating performance and market-based valuation measures to compare firms that are *best-in-class* or *worst-in-class* in terms of CSR in a given industry to the remaining firms in the industry. While many past studies of the relation between CSR and financial performance have relied upon ROA (defined as Net Income / Assets) to measure financial performance, we contend that operating performance (defined as Operating Income Before Depreciation (OIBD) / Total Assets) is the appropriate measure because it does not penalize firms for use of leverage in their capital structure and excludes the effects of taxation (Barber and Lyon, 1996; Loughran and Ritter, 1997). To compare relative financial market valuations, we use an estimate of Tobin's Q, calculated as [(Total Assets – Book Value of Common Equity + Market Value of Common Equity) / Total Assets].

The use of ratios can lead to problematic inferences when extreme values occur (Certo *et al.*, 2018; Henderson, Raynor, and Ahmed, 2012), and therefore we take several steps to alleviate this concern. First, we require all firms in the sample to have at least $50 million in sales and $100 million in assets throughout the sample period. This requirement reduces the influence of extreme values in ratios of operating performance and/or Tobin's Q from firms that are extremely small in terms of sales or assets and/or lack profitability due to the early-stage investment requirements of the firm (e.g., early-stage R&D-intensive Biotech firms).[2] Second, we construct financial performance ratios at an aggregate level for each industry and year and we report univariate parametric and nonparametric comparisons of *best-in-class* firms with the other firms in the industry, including *worst-in-class* firms. Third, we test whether *best-in-class* CSR firms financially outperform their competitors using indicator variables and estimating multivariate regressions with industry by year fixed effects and robust clustered standard errors (Gormley and Matsa, 2014), including control variables that have been associated with our measures of financial performance. Finally, we implement a two-stage least squares (2SLS) instrumental variables approach to control for endogeneity using geographic-based instruments that take into account the concentration of other local firms that are *best-in-class*. Following Karpoff *et al*. (2017), we define local firms to be all firms headquartered within a 100-mile radius of the focal firm. Our first instrument is the percentage of local firms in other industries that are *best-in-class*. As a second instrument, we use the average within-industry percentile CSR ranking of firms headquartered in the same state.

Sample

To examine firm CSR performance, we use firm-level ratings from the MSCI ESG Stats Database, formerly known as the Kinder, Lydenberg and Domini database (KLD). The database provides binary ratings (0/1) for each firm based on its evaluation of the presence and/or absence of strengths and concerns for particular indicators related to seven CSR dimensions including employee relations, community, product, corporate governance, diversity, environment, and

[2] Including small firms with little or no assets or sales in our study would result in extreme values for several financial ratios important to our analysis. We contend that the best approach is to exclude these firms from the analysis from the start (see internet Appendix A and corresponding Tables A1 and A2).

human rights. KLD uses analysts and publicly available information from databases, financial reports, and the popular press to rate firms on the various dimensions. While other metrics have been used to assess social responsibility, the KLD score is a good overall assessment of a firm's involvement in CSR activities and has been widely used in the CSR literature (Barnett and Salomon, 2012; Servaes and Tamayo, 2013; Shiu and Yang, 2017). To rank firms' CSR activities within a given year and industry, we follow previous research and use the firm's overall or cumulative KLD score, which consists of the sum of the positive attributes of the firm on each of the dimensions less the sum of the negative attributes or concerns (Barnett and Solomon, 2012).[3]

Our sample consists of firms headquartered in the United States for which we have KLD scores and corresponding financial data from Compustat.

We begin our analysis in 2003 and end it in 2013. 2003 corresponds to when KLD covered the approximately 2,400 firms included in the MSCI Investable Market Index, constituting around 99 percent of the market capitalization of tradable equities in the United States. The population of firms with available data prior to 2003 is a fraction of this population. After 2013 there is a substantial dropoff in data items/subcategories in the KLD database, and the amount of missing values is significant, making the data in 2014 and 2015 incompatible with the 2003 to 2013 data. Consequently, the coverage is relatively stable over the period that we study, ensuring that our results are not influenced by a significant change in sample composition over time. Given that the total number of individual strengths and concerns tracked by KLD changes during the sample period, we use cumulative KLD scores to rank firms relative to their industry peers each year. In order to address the lack of stability and consistency in the time series, we focus on this rank variable in our empirical tests to ensure that our results are not affected by the changing total

[3] We find similar results if we exclude the corporate governance components from aggregate KLD scores.

number of KLD attributes in our cross-sectional and time-series analysis. We require a minimum of 10 firms in each industry-year to identify the top and bottom decile of firms within a given industry-year. There are 31 FFI48 classifications that meet these criteria, resulting in a total of 23,369 firm-year observations across the 2003-2013 sample period and 341 industry by year observations.

----------Insert Table 1 here------------

Table 1 reports the distribution of cumulative KLD scores from 2003 to 2013. The overall mean KLD scores decrease from 2003 until 2011 and then increase by nearly 2 points from 2011 to 2012. When looking at minimums and maximums, the minimum KLD score in 2003 is -11, whereas by 2013 it is -6. Similarly, the maximum KLD score increases from 9 in 2003 to 17 in 2013. Because of the positive drift in KLD scores over time, a simple comparison of firms with high KLD scores to firms with low KLD scores without regard for time would be problematic because of instability in the scores across the sample period.

Table 2 provides the distribution of Cumulative KLD scores for the sample across the 31 FFI48 classifications with sufficient firms to meet our sample criteria. KLD scores differ significantly across industries, with certain segments exhibiting lower average scores (e.g., *Construction* and *Petroleum and Natural Gas*) and other industries having much higher average KLD scores (e.g., *Consumer Goods*). Simple comparisons of firms with high KLD scores to firms with low KLD scores without industry controls would essentially contrast firms from different industries. In addition, distributions within individual industries show significant variation in KLD scores, with a range for *Electronic Equipment* of 27 (low of -8 to high of 19) but only 10 for *Personal Services* (low of -6 and a high of 4). To overcome these problems and time series issues raised above, we contend that a valid research design when studying the relation between KLD ratings and corporate performance must incorporate within-year and within-industry comparison of firms.

-----------Insert Table 2 here-----------

Given the variation in KLD scores across industries, it is also important to recognize that a similar-sized deviation from the mean or median may not be comparable across industries with a wide range of KLD scores and industries with a relatively small range in KLD scores. In addition, clustering around the median for the KLD cumulative score within an industry affects the ability to separate a firm's CSR performance from its peers. Using a 10 percent threshold ensures that the firms that are classified as *best-in-class* or *worst-in-class* are truly different from the median firm in the industry and from each other, and that our analysis focuses on firms with substantially different CSR performance as opposed to a comparison with an average firm (see internet Appendix B and corresponding Figures B1 and B2 for further discussion and examples).

Because of the drift in cumulative KLD scores over time and industry clustering, we focus our subsequent tests on comparing firms that appear to be *best-in-class* with the remaining firms including the *worst-in-class* firms in any given year. To do so, we segment the data by both industry and year into three groups of firms, consisting of the top 10 percent (*best-in-class*), the bottom 10 percent (*worst-in-class)*, and the remaining 80 percent or middle firms (*mid-80%*). To ensure a balanced panel across the *best-in-class* and *worst-in-class* subsamples, we break any ties at the breakpoint for the lower or upper decile by randomly assigning firms that are tied with the same breakpoint KLD cumulative score to the middle group and the upper or lower decile.[4]

Results

Univariate Comparisons

Table 3 presents comparisons of operating performance and Tobin's Q within the 341 industry-year classifications. Operating performance is a commonly used measure of profitability because it excludes the effects of capital structure choices and taxation (Barber and Lyon, 1996; Loughran and Ritter, 1997). Tobin's Q is a commonly used measure of value-added in the economics, finance and management literature (Morck, Shleifer, and Vishny, 1988; Servaes and Tamayo, 2013).

We include comparisons between *best-in-class, mid-80%*, and *worst-in-class firms.* If CSR activities are truly related to financial performance, it should be more apparent with the contrasts of the *best-in-class* and *worst-in-class* firms. Because both of our dependent variables (operating performance and Tobin's Q) are ratios where mean values can be materially influenced by outliers, we calculate the ratios at the aggregate level for our comparisons in Table 3. Specifically, in terms of operating performance, for all firms in a given industry, year, and classification (*best-in-class; mid-80%; worst-in-class*), we sum the operating profit (operating income before depreciation) for all firms and then divide the total by the sum of the assets for all firms. Thus, we have a single measure of profitability for each comparison group without any distortions that may arise from individual outlier ratios. We perform a similar calculation for our measure of Tobin's Q, summing

[4] Our procedure for random assignment consists of sorting the firms first by KLD cumulative score and second by a randomly assigned identifier for each firm-year using the *runiform* function in Stata. We also truncate the number of firms classified as *best-in-class* and *worst-in-class* as no more than 10% of the firms in that industry-year, meaning that the number of firms so classified is somewhere between 9% and 10% for each industry-year.

up the market value of assets for the entire group and then dividing by the sum of the book value of assets for that group. We report medians as well as means because Barber and Lyon (1996) note that nonparametric tests such as Wilcoxon signed ranks test are more powerful than parametric tests because of the influence of extreme observations that can show up when analyzing ratios.

----------Insert Table 3 here----------

Panel A of Table 3 presents the operating performance comparisons. Firms within the bottom 10 percent of KLD scores for their industry in a given year have a mean and median aggregate operating performance of 11.36 percent and 11.30 percent, respectively. This compares to a mean and median of 14.19 percent and 14.57 percent for firms in the top 10 percent of KLD scores in the industry in a given year, and both the mean and median differences are significant ($p = 0.000$). *Best-in-class* firms also outperform the middle 80 percent of firms ($p = 0.000$), showing that *best-in-class* firms have higher levels of operating performance than industry peers.

Panel B of Table 3 presents the Tobin's Q comparisons. The mean and median levels of Tobin's Q for the *best-in-class* firms in an industry for a given year are significantly higher than the Tobin's Q for the *worst-in-class* firms and the middle 80 percent of firm ($p < 0.01$). Collectively, these results suggest that the *best-in-class* firms not only outperform the *worst-in-class* and other firms from an accounting perspective, but also that markets value the *best-in-class* firms more highly.

Multivariate Analysis

While the univariate results are compelling, we now turn to a multivariate setting with industry by year fixed effects and include control variables known to correlate with financial performance.

We present a correlation matrix in Table 4 for the variables used in our multivariate tests in Table 5.

----------Insert Table 4 here-----------

Model 1 of Table 5 uses operating performance (OIBD/TA) as the dependent variable, and Models 2 and 3 use Tobin's Q as the dependent variable.[5] The models utilize the same independent control variables, including the natural log of total assets, the debt ratio, the natural log of the number of employees, Advertising expenses as a percentage of sales, R&D expenses as a percentage of sales, and Selling, General, & Administrative expenses as percentage of sales. The models also include indicator variables to test whether our univariate findings are robust to a multivariate setting. We have three categories of CSR scores: the *worst-in-class* or bottom 10 percent of the industry in a given year, the middle 80 (*mid-80%* CSR) percent, and the *best-in-class* or top 10 percent of the industry in a given year. Our model, therefore, includes two dummy variables, one for the *mid-80% CSR* and one for the *best-in-class CSR*, with the *worst-in-class* firms representing the base model. Due to both industry clustering and time trends in KLD scores in the sample, we estimate our model including industry by year fixed effects and utilize robust standard errors clustered by industry-year. By estimating regressions controlling for industry by year fixed effects, we are analyzing the variations around the mean for each observation given the industry-year combination. All continuous variables are winsorized at the 1 percent and 99 percent levels to mitigate the potential influence of extreme outliers.

----------Insert Table 5 here----------

[5] Unlike the univariate comparisons shown in Table 3 where the performance measures are aggregated by industry and year, the performance measures in the multivariate analysis are firm and year specific.

Our primary interest in Model 1 lies in whether the *best-in-class* CSR firms have higher levels of operating performance in a multivariate setting with control variables. The coefficient on the *best-in-class* CSR variable is 2.2 percent ($p = 0.000$) and suggests that, controlling for year, industry, and other variables, firms at the upper end of their industry in terms of CSR scores experience significantly higher levels of operating performance than *worst-in-class* firms, which are the base model. The coefficient on *mid-80% CSR* is also positive and significant indicating outperformance relative to the *worst-in-class* firms, but the coefficient is of lower magnitude than *best-in-class* and a statistical test of the difference in the coefficients ($p <0.01$) indicates that the *best-in-class* firms also have greater operating performance than the middle 80 percent firms. We obtain similar results in Model 2 that focuses on Tobin's Q. Tobin's Q ratios are higher for *best-in-class* firms than *worst-in-class* ($p = 0.000$) and the remaining firms[6], indicating the results are robust to a multivariate setting including control variables, fixed effects and robust clustered standard errors.

The results presented in Models 1 and 2 of Table 5 confirm our univariate findings that *best-in-class* CSR firms have higher levels of operating performance and higher relative market valuations. However, firms earning higher operating profits generally have more cash flows at their disposal to invest in CSR. These higher profits could lead to higher relative valuations (Tobin's Q), even if at the margin *best-in-class* firms are overspending on CSR activities and harming shareholder wealth. This possibility is central to the debate regarding whether CSR actually creates shareholder value or whether higher profitability *enables* managers to divert extra resources toward CSR and "do good" simply because they are "doing well." In order to consider

[6] Statistical tests of the difference between the best-in-class and mid-80% coefficients confirm that the coefficients significantly differ ($p < 0.01$).

this possibility without the potential omitted variable bias, we re-estimate the model using Tobin's Q as the dependent variable and include our measure of operating performance as a control variable (Model 3). Relative to Model 2, the explanatory power of Model 3 increases significantly, with the R-squared increasing to 40.39 percent from 17.52 percent. The coefficient on OIBD/TA is positive ($p = 0.000$) and the coefficient on *best-in-class* CSR is smaller at 0.114, yet retains its statistical significance ($p = 0.000$). This result suggests that *best-in-class* firms experience higher relative valuations, even after controlling for profitability. Hence, we interpret our results as evidence that *best-in-class* CSR firms, all else equal, receive higher relative valuations in financial markets and CSR leaders within an industry are not simply doing good because they are doing well.

Addressing endogeneity concerns with an instrumental variable approach

Despite the inclusion of higher order (industry by year) fixed effects to control for unobservable industry-year level heterogeneity, it is possible that an omitted endogenous firm characteristic correlated with both CSR rank and operating performance/market valuation could drive our results. We implement a two-stage least squares (2SLS) instrumental variables approach to address this endogeneity concern. A valid instrument for the endogenous variable (CSR rank) must meet two conditions: the relevance condition and exclusion restriction. The relevance condition requires a non-zero correlation between the endogenous variable and the instrument. The exclusion restriction requires that the instrument is only indirectly related to the outcome variable (operating performance and/or relative market valuation) through its effect on the endogenous variable.

Husted *et al*. (2016) show that geography plays a significant role in CSR activities – specifically, firms located near large cities tend to have higher levels of CSR. Thus, a strong case can be made for using the CSR activities of local firms as instrumental variables to allay potential endogeneity concerns. We employ two geographic-based instruments, with the first capturing the concentration of local firms that are *best-in-class* and the second capturing the average within-industry-year percentile CSR ranking for all other firms headquartered in the same state as the focal firm. We follow the approach of Karpoff *et al*. (2017) to construct our local instrument. For each firm-year, we find all companies headquartered within a 100-mile radius of each sample firm based on the firm's headquarters zip code. We exclude any companies in the same Fama-French 48 industry classification or that are located in a different state. If we find fewer than three local matches, we expand the 100-mile radius to include the entire state[7]. The number of firms that are *best-in-class* CSR within their respective industries are then scaled by the total number of local firms. Our local instrument is the percentage of non-industry peer firms headquartered within the 100-mile radius that are *best-in-class* CSR. As noted earlier, our second instrument is the average within-industry-year percentile CSR ranking for all other firms headquartered in the same state as the focal firm. Given the evidence from Husted *et al*. (2016) establishing a link between the geographic region of a firm's headquarters and CSR scores, the instruments are plausibly correlated with our CSR classifications. The instruments also reasonably meet the exclusion restriction, as the decision regarding headquarters location is generally determined early in a firm's life and is unlikely to be related to the firm's current performance or relative market valuation. In addition, the relative CSR rankings of local companies from different industries are unlikely to have an impact on a firm's relative performance within its own industry-year.

----------Insert Table 6 here----------

[7] There are 287 (50) firm-years (firms) out of 23,369 (3,513) total firm-years (firms) that have less than 3 local matches within the 100 mile radius.

Table 6 summarizes the results of the two-stage least squares analysis. The first stage model regresses the endogenous indicator variable on both instruments and all controls from Table 5 to estimate the predicted value. In the second stage, the endogenous variable (*best-in-class* CSR) is replaced with its predicted value (instrumented endogenous variable) from the respective first stage models. Operating performance and Tobin's Q are then regressed on the instrumented endogenous variable and the same controls from the first stage.

Models 1 and 3 of Table 6 report the first stage regressions estimating the likelihood that a firm is *best-in-class* CSR in a given industry-year, incorporating the variables used in the second stage models (Model 1 is the first stage for Model 2 and Model 3 is the first stage for Model 4). The results suggest that the instruments are significant determinants of CSR rankings, as the F-statistics (63.18 in Model 1; 58.35 in Model 3) are higher than the critical value of 16.38 proposed by Stock and Yogo (2005). Models 2 and 4 report the second stage regressions estimating the effect of instrumented (predicted values from first stage) *best-in-class* CSR rankings on operating performance (Model 2) and Tobin's Q (Model 4). Models 2 and 4 report the second stage regressions estimating the effect of instrumented (predicted values from first stage that are a continuous measure capturing the probability of being best-in-class) best-in-class CSR rankings on operating performance (Model 2) and Tobin's Q (Model 4). We test the null hypothesis that the estimates provided by ordinary least squares (OLS) and two-stage least squares (2SLS) only differ due to sampling error. If the test statistic is significant, we need to treat our variables of interest (i.e. *best-in-class*) as endogenous. The F-tests for endogeneity support the use of 2SLS estimates to draw conclusions regarding the relations between *best-in-class* CSR, operating performance and firm market valuation.

After controlling for endogeneity, the significant positive relation between operating performance and *best-in-class* CSR rankings shown in the earlier univariate and multivariate tests disappears – the coefficient on the probability of being *best-in-class* (Model 2) is insignificant (p = 0.209). The relation between *best-in-class* CSR ranking and Tobin's Q (Model 4), however, remains positive (p = 0.000). Two instruments allow us to perform several overidentification diagnostic tests. A statistically significant result in these tests suggests that the instruments used may not be valid. The p-values for the Sargan (Basmann) test statistic in Models 2 and 4 are 0.0790 (0.0810) and 0.2173 (0.2207), respectively. Even after controlling for endogeneity, there remains a positive and significant relationship between relative market valuation (Tobin's Q) and *best-in-class* CSR status. The economic effect of *best-in-class* CSR ranking on Tobin's Q is also meaningful: a one standard deviation increase in the probability of being *best-in-class* is associated with an increase in Tobin's Q of 0.223.

Table 7 provides an additional test where we estimate the same models in Table 6, but using a Limited Information Maximum Likelihood (LIML) approach. As discussed in Stock and Yogo (2005) and Angrist and Pischke (2009), relative to 2SLS, LIML is more robust against possible bias from weak instruments. Angrist and Pischke (2009) recommend that researchers estimate LIML models as robustness checks when using over-identified 2SLS models.

----------Insert Table 7 here----------

Across both 2nd stage models (Models 2 and 4) the coefficients and diagnostic test statistics are virtually identical to those obtained using the 2SLS approach in Table 6, providing further evidence regarding the validity of our instruments. In terms of economic significance, the relation between a *best-in-class* CSR ranking and Tobin's Q is nearly the same as estimated from the 2SLS regression: a one standard deviation increase in the probability of being *best-in-class* is associated with an increase in Tobin's Q of 0.227.

Additional Robustness Tests

We perform additional tests to explore the robustness of the instrumental variable analysis. First, we use a 50-mile radius, rather than a 100-mile radius, to construct our instrument and find consistent results (see internet Appendix C and corresponding Table C3). Second, we obtain similar results when we exclude regulated firms (i.e., financial firms and utilities) (See internet Appendix C and corresponding Table C4). In sum, we find consistent results using both 2SLS and LIML estimations and in all of our additional robustness tests.

Discussion and Conclusion

When we began our investigation into CSR and firm financial performance, our initial plans were to examine a sample of firms that substantially increased their CSR scores relative to industry peers and then track those firms to see if financial performance materially changed. In other words, our aim was to find firms that transitioned from *worst-in-class* to *best-in-class*, resulting from significant additional investment in CSR. Given the long history of CSR scores in the KLD data, we were optimistic that we could identify a sufficiently large sample that made a significant additional investment in CSR. As we became more familiar with the distribution of CSR ratings in KLD, the trends over time, the significant differences across industries, and the clustering around the median within industries, it became apparent that transitions from *worst-in-class* to *best-in-class* are so infrequent that such an analysis is simply not feasible. Following Henderson, *et al*. (2012), one avenue for future research would be to develop a better understanding of those firms with sustained *best-in-class* CSR performance.

These issues may partially explain why studies of additions and/or deletions from indices consisting of *best-in-class* CSR such as Ramchander *et al.* (2012), Hawn *et al.* (2018), and Durand *et al*. (2019) find mixed or insignificant market reactions. For example, Hawn *et al.* (2018) and Durand *et al.* (2019) examine the Dow Jones Sustainability Index (DJSI), which includes firms only if they are among the top 10 percent CSR leaders in the industry. Given the distributional properties of CSR ratings, at least in the CSR scores that we examine, it is unlikely that the addition to or deletion from an index that tracks *best-in-class* within an industry reflects a material shift in the CSR commitment of the covered firms. In other words, given the level of disclosure of CSR-related activities by firms, investors likely know that firms that are covered by, added to, or deleted from the DJSI index are all among the top firms in their given industry and that the re-composition of the index to include or exclude a particular firm does not materially affect investor opinions of a firm's commitment to CSR. While confirming the muted stock price reactions found by Hawn *et al.* (2018), Durand *et al.* (2019) uncover another channel through which *best-in-class* status may influence overall company valuations. Specifically, they report that *best-in-class* recognition through inclusion in the DJSI leads to higher levels of analyst coverage and continuation on the DJSI leads to an increase in the proportion of shares held by longer-term investors. Given that Chung and Jo (1996) report a positive relation between analyst coverage and Tobin's Q and Luo *et al*. (2015) find that analysts reduce information asymmetry related to CSR between firms and investors, the increased coverage and visibility associated with being recognized as *best-in-class* may partially explain the robustness of our finding that *best-in-class* firms have higher relative valuations.

By contrasting the *best-in-class* or top 10 percent within an industry to other firms in the industry including the *worst-in-class*, we believe our research design more accurately distinguishes between firms that make significant investment in CSR relative to those that do not. This approach also lessens the likelihood that the distributional properties of CSR scores and/or financial ratios often used to study firm performance skew our results. We find that *best-in-class* CSR firms have higher levels of operating performance and higher relative valuations (i.e., Tobin's Q) than their *worst-in-class* and *mid-80%* counterparts in both univariate tests and in multivariate tests that include control variables known to correlate with financial performance measures. The multivariate tests utilize industry by year fixed effects and robust clustered standard errors which have been shown by Gormley and Matsa (2014) to be the best method to obtain consistent estimates in the presence of unobserved group heterogeneity. When we adopt an instrumental variables approach in order to mitigate potential endogeneity concerns, the significantly higher levels of operating performance for *best-in-class* firms do not persist, which calls into question whether the relation between CSR and operating performance is causal. This suggests that endogeneity issues may partially explain why past studies have found mixed results when examining the relation between CSR and operating performance. We continue to find that *best-in-class* firms have significantly higher relative market valuations (Tobin's Q).

Our research provides managers and investors with important implications including the value of relative benchmarking. Specifically, managers and investors need to understand that expectations of CSR performance evolve over time and that when investors evaluate a firm's CSR relative their industry peers, they place higher valuations on *best-in-class* firms within the industry.

TABLE 1
Distribution of Cumulative KLD Scores by Year

The sample includes firms headquartered in the United States in the Fama-French 48 industries with at least 10 firms in each year over the 2003 – 2013 sample period. Firms with limited sales (less than $50 million) and/or limited assets (less than $100 million) at any time during the sample period are excluded. The final sample consists of 23,369 firm-year observations.

Year	Count	Mean KLD score	Minimum	10th percentile	25th percentile	Median	75th percentile	90th percentile	Maximum
2003	2,283	-0.21	-11	-2	-1	0	1	2	9
2004	2,302	-0.52	-10	-3	-2	-1	0	2	10
2005	2,284	-0.54	-9	-3	-2	-1	0	2	11
2006	2,240	-0.62	-10	-3	-2	-1	0	2	14
2007	2,158	-0.60	-11	-3	-2	-1	1	2	15
2008	2,143	-0.58	-11	-3	-2	-1	1	2	14
2009	2,108	-0.59	-11	-3	-2	-1	1	2	14
2010	2,081	-0.59	-8	-3	-2	-1	0	3	17
2011	1,997	-0.91	-8	-3	-3	-2	0	4	19
2012	2,000	0.83	-6	-1	-1	0	1	4	16
2013	1,773	0.80	-6	-2	-1	0	2	5	17
Total	23,369	-0.34	-11	-3	-2	-1	1	2	19

TABLE 2
Distribution of Cumulative KLD Scores by Industry Classification

The sample includes firms headquartered in the United States in the Fama-French 48 industries with at least 10 firms in each year over the 2003 – 2013 sample period. Firms with limited sales (less than $50 million) and/or limited assets (less than $100 million) at any time during the sample period are excluded. The final sample consists of 23,369 firm-year observations.

Industry Classification	N	Mean KLD score	Minimum	10th percentile	25th percentile	Median	75th percentile	90th percentile	Maximum
Food Products	422	0.12	-9	-4	-2	0	1	5	17
Entertainment	344	-1.01	-7	-3	-2	-1	0	1	6
Printing and Publishing	219	0.35	-6	-2	-1	0	2	3	7
Consumer Goods	446	0.77	-4	-2	-1	0	2	4	16
Apparel	342	-0.21	-8	-3	-2	-1	1	3	13
Healthcare	398	-1.06	-6	-3	-2	-1	0	1	7
Medical Equipment	531	-0.35	-4	-2	-1	-1	0	2	9
Pharmaceutical Products	537	0.41	-8	-3	-2	0	1	5	15
Chemicals	582	-1.36	-8	-4	-3	-1	0	2	7
Construction Materials	400	-0.63	-6	-3	-2	-1	0	2	9
Construction	374	-1.44	-11	-4	-2	-1	0	1	5
Steel Works	347	-1.13	-8	-4	-2	-1	0	1	12
Machinery	837	-0.65	-7	-3	-2	-1	0	2	11
Electrical Equipment	304	-0.69	-8	-3	-2	-1	0	1	7
Automobiles and Trucks	376	-1.00	-8	-3	-2	-1	0	1	9
Petroleum and Natural Gas	1,053	-1.45	-11	-4	-3	-1	0	1	11
Utilities	947	-0.32	-10	-4	-2	0	1	4	15
Communication	819	-0.52	-7	-3	-2	-1	0	2	9
Personal Services	373	-0.81	-6	-3	-2	-1	0	1	4
Business Services	2,662	-0.25	-8	-3	-2	0	1	2	15
Computers	835	0.24	-5	-2	-2	0	1	4	14
Electronic Equipment	1,382	0.05	-8	-3	-2	-1	1	3	19
Measuring and Control Equipment	457	0.06	-3	-2	-1	0	1	3	14
Business Supplies	316	0.34	-7	-3	-1	0	2	5	10
Transportation	730	-0.70	-7	-3	-2	-1	0	1	9
Wholesale	785	-0.72	-6	-3	-2	-1	0	1	9
Retail	1,582	-0.48	-10	-3	-2	-1	1	2	13
Restaurants, Hotels, Motels	426	-0.34	-8	-3	-2	-1	0	2	13
Banking	2,636	0.36	-5	-2	-1	0	1	3	14
Insurance	1,060	-0.32	-7	-3	-2	0	1	2	10
Trading	847	-0.53	-8	-3	-2	-1	0	2	8
Full Sample	23,369	-0.34	-11	-3	-2	-1	1	2	19

TABLE 3
Comparisons of Operating Performance and Tobin's Q Across Industry and Year KLD classifications

Operating performance is measured as Operating Income Before Depreciation scaled by Total Assets (OIBD / TA) is calculated at the aggregate level for each of the 341 observations, consisting of 11 years and 31 industries sufficiently large to conduct our analyses. Specifically, for a given industry, year, and decile classification, Operating Income Before Depreciation is summed across all firms and then scaled by the sum of Total Assets. Tobin's Q is calculated as [(Total Assets – Book Value of Common Equity + Market Value of Common Equity) / Total Assets], and is aggregated in a similar fashion, where the market value of assets are summed up and then scaled by the sum of asset book values. *Worst-in-class* firms and *best-in-class* firms are firms ranked by cumulative KLD CSR score in the bottom and top deciles, respectively, for the industry and year. P-values are in parentheses.

		worst-in-class KLD score (1)	KLD score in the middle 80% (2)	*best-in-class* KLD score (3)	(2-1)	(3-2)	(3-1)
PANEL A							
OIBD / TA	Mean	0.1136	0.1238	0.1419	0.0102	0.0181	0.0283
					(0.053)	(0.000)	(0.000)
	Median	0.1130	0.1287	0.1457	0.0157	0.0170	0.0327
					(0.000)	(0.000)	(0.000)
PANEL B							
Tobin's Q	Mean	1.5739	1.6457	1.8700	0.0718	0.2243	0.2961
					(0.242)	(0.000)	(0.000)
	Median	1.4580	1.5487	1.7380	0.0907	0.1893	0.2800
					(0.056)	(0.001)	(0.000)

TABLE 4
Correlation Matrix

Variable	Tobin's Q	OIBD / TA	Natural log of Assets	Debt Ratio	Natural log of # of Employees	Advertising Expense / Sales	R&D Expense / Sales	SG&A Expense / Sales	*worst-in-class* CSR	*mid-80%* CSR	*best-in-class* CSR	Cumulative KLD Score
Tobin's Q	1											
OIBD / TA	0.450	1										
Natural log of Assets	-0.234	-0.124	1									
Debt Ratio	-0.199	-0.162	0.404	1								
Natural log of # Employees	-0.027	0.200	0.576	0.109	1							
Advertising Expense / Sales	0.129	0.062	-0.036	0.010	0.051	1						
R&D Expense / Sales	0.224	-0.181	-0.159	-0.224	-0.151	-0.014	1					
SG&A Expense / Sales	0.243	-0.155	-0.248	-0.181	-0.169	0.266	0.538	1				
worst-in-class CSR	-0.029	-0.022	0.067	0.042	0.077	-0.014	-0.013	-0.047	1			
mid-80% CSR	-0.011	-0.013	-0.232	-0.048	-0.222	-0.006	0.002	0.031	-0.670	1		
best-in-class CSR	0.043	0.039	0.244	0.022	0.220	0.022	0.010	0.005	-0.103	-0.670	1	
Cumulative KLD Score	0.071	0.021	0.230	0.031	0.175	0.060	0.047	0.092	-0.413	-0.142	0.603	1
Mean	1.774	0.116	7.545	0.575	1.703	0.012	0.036	0.225	0.093	0.813	0.093	-0.345
Median	1.400	0.110	7.365	0.563	1.443	0.000	0.000	0.200	0	1	0	-1
Standard Deviation	1.134	0.101	1.621	0.274	1.231	0.029	0.101	0.197	0.291	0.390	0.291	2.481
Minimum	0.347	-1.691	4.606	0	0	0	0	0	0	0	0	-11
Maximum	20.274	1.389	14.698	7.712	7.697	0.489	3.938	5.216	1	1	1	19

TABLE 5
Multivariate Analysis

The dependent variable in Model 1 is Operating Income Before Depreciation / Total Assets (OIBD/TA). The dependent variable in Models 2 and 3 is Tobin's Q, calculated as [(Total Assets – Book Value of Common Equity + Market Value of Common Equity) / Total Assets]. Assets is the natural log of total assets. Debt ratio is total liabilities / total assets. Employees is the natural log of the number of employees. Advertising Expense, R&D Expense, and SG&A Expense are expressed as a percentage of sales and missing values are set to zero. *Mid-80%* CSR is an indicator variable equal to one (zero otherwise) for firms that are not in either the lowest or highest decile of KLD scores for their Fama-French 48 industry classification in a given year. *Best-in-class* CSR is an indicator variable equal to one (zero otherwise) for firms in the top decile of KLD scores for their Fama-French 48 industry classification in a given year. All models include Fama-French 48 industry by year fixed effects and robust standard errors clustered at the industry by year level. All variables are winsorized at the 1 percent and 99 percent levels to mitigate the potential influence of outliers. P-values are in parentheses.

Model	(1)	(2)	(3)
	OIBD/TA	Tobin's Q	Tobin's Q
Intercept	0.235	2.658	1.157
	(0.000)	(0.000)	(0.000)
Assets	-0.014	-0.152	-0.059
	(0.000)	(0.000)	(0.000)
Debt Ratio	-0.060	-0.343	0.039
	(0.000)	(0.000)	(0.244)
Employees	0.022	0.123	-0.016
	(0.000)	(0.000)	(0.155)
Advertising Expense	0.260	3.830	2.172
	(0.000)	(0.000)	(0.000)
R&D Expense	-0.249	1.754	3.341
	(0.000)	(0.000)	(0.000)
SG&A Expense	-0.066	0.395	0.819
	(0.000)	(0.003)	(0.000)
OIBD/TA			6.386
			(0.000)
***mid-80%* CSR**	0.007	0.021	-0.026
	(0.000)	(0.277)	(0.173)
***best-in-class* CSR**	0.022	0.253	0.114
	(0.000)	(0.000)	(0.000)
Industry by Year Fixed Effects	Yes	Yes	Yes
N	23,369	23,369	23,369
F-Statistic	120.05	47.53	166.43
R-Squared	21.67%	17.52%	40.39%

TABLE 6
Instrumental Variable Approach

The dependent variable in the first stage models is an indicator for *best-in-class* CSR. The instrumental variables in the first stage models are the percentage of local firms (headquartered within a 100-mile radius) in other industries that are *best-in-class* CSR and the average within-industry-year percentile CSR ranking for all other firms headquartered in the same state as the focal firm. Models 2 and 4 show the second stage of the two-stage least squares regressions, where the dependent variable in Model 2 is Operating Income Before Depreciation / Total Assets (OIBD/TA), and the dependent variable in Model 4 is Tobin's Q, calculated as [(Total Assets – Book Value of Common Equity + Market Value of Common Equity) / Total Assets]. Assets is the natural log of total assets. Debt ratio is total liabilities / total assets. Employees is the natural log of the number of employees. Advertising Expense, R&D Expense, and SG&A Expense are expressed as a percentage of sales and missing values are set to zero. *Best-in-class* CSR is the fitted value from the first stage Model 1 or Model 3 and captures the probability that a firm is in the top decile of KLD scores for their Fama-French 48 industry classification in a given year. All models include Fama-French 48 industry by year fixed effects and robust standard errors clustered at the industry by year level. All variables are winsorized at the 1 percent and 99 percent levels to mitigate the potential influence of outliers. P-values are in parentheses.

	1st stage	2nd stage	1st stage	2nd stage
Model	(1)	(2)	(3)	(4)
	best-in-class CSR	OIBD/TA	best-in-class CSR	Tobin's Q
Intercept	-0.361	0.235	-0.419	1.909
	(0.000)	(0.000)	(0.000)	(0.000)
Assets	0.041	-0.012	0.044	-0.165
	(0.000)	(0.000)	(0.000)	(0.000)
Debt Ratio	-0.069	-0.064	-0.055	0.173
	(0.000)	(0.000)	(0.000)	(0.000)
Employees	0.031	0.023	0.026	-0.078
	(0.000)	(0.000)	(0.000)	(0.000)
Advertising Expense	0.295	0.278	0.235	1.540
	(0.002)	(0.000)	(0.013)	(0.000)
R&D Expense	0.228	-0.235	0.283	2.591
	(0.000)	(0.000)	(0.000)	(0.000)
SG&A Expense	0.071	-0.061	0.085	0.609
	(0.000)	(0.000)	(0.000)	(0.000)
OIBD/TA			0.224	5.854
			(0.000)	(0.000)
***best-in-class* CSR**		-0.042		2.516
		(0.209)		(0.000)
Percent of nearby firms that are *best-in-class* CSR	0.200		0.199	
	(0.000)		(0.000)	
Avg. percentile CSR ranking of firms in same state	0.157		0.163	
	(0.001)		(0.001)	
N	23,369	23,369	23,369	23,369
F-Statistic	63.18		58.35	
Wald Chi-squared		1,114.48		11,490.61
F-Test for Endogeneity, p-value		0.0645		0.0000
Sargan overidentification test, p-value		0.0790		0.2173
Basmann overidentification test, p-value		0.0812		0.2207
R-squared	8.75%	29.44%	9.07%	40.50%

TABLE 7
Instrumental Variable Approach using a LIML Estimation Model

The dependent variable in the first stage models is an indicator for *best-in-class* CSR. The instrumental variables in the first stage models are the percentage of local firms (headquartered within a 100-mile radius) in other industries that are *best-in-class* CSR and the average within-industry-year percentile CSR ranking for all other firms headquartered in the same state as the focal firm. Models 2 and 4 show the second stage of the limited-information maximum likelihood (LIML) regressions, where the dependent variable in Model (2) is Operating Income Before Depreciation / Total Assets (OIBD/TA), and the dependent variable in Model (4) is Tobin's Q, calculated as [(Total Assets – Book Value of Common Equity + Market Value of Common Equity) / Total Assets]. Assets is the natural log of total assets. Debt ratio is total liabilities / total assets. Employees is the natural log of the number of employees. Advertising Expense, R&D Expense, and SG&A Expense are expressed as a percentage of sales and missing values are set to zero. *Best-in-class* CSR is the fitted value from the first stage Model 1 or Model 3 and captures the probability that a firm is in the top decile of KLD scores for their Fama-French 48 industry classification in a given year. All models include Fama-French 48 industry by year fixed effects and robust standard errors clustered at the industry by year level. All variables are winsorized at the 1 percent and 99 percent levels to mitigate the potential influence of outliers. P-values are in parentheses.

	1st stage	2nd stage	1st stage	2nd stage
Model	(1)	(2)	(3)	(4)
	best-in-class CSR	OIBD/TA	best-in-class CSR	Tobin's Q
Intercept	-0.361	0.235	-0.419	1.925
	(0.000)	(0.000)	(0.000)	(0.000)
Assets	0.041	-0.012	0.044	-0.167
	(0.000)	(0.000)	(0.000)	(0.000)
Debt Ratio	-0.069	-0.064	-0.055	0.175
	(0.000)	(0.000)	(0.000)	(0.000)
Employees	0.031	0.024	0.026	-0.079
	(0.000)	(0.000)	(0.000)	(0.000)
Advertising Expense	0.295	0.279	0.235	1.527
	(0.002)	(0.000)	(0.013)	(0.001)
R&D Expense	0.228	-0.234	0.283	2.576
	(0.000)	(0.000)	(0.000)	(0.000)
SG&A Expense	0.071	-0.061	0.085	0.605
	(0.000)	(0.000)	(0.000)	(0.000)
OIBD/TA			0.224	5.844
			(0.000)	(0.000)
***best-in-class* CSR**		-0.044		2.565
		(0.202)		(0.000)
Percent of nearby firms that are *best-in-class* CSR	0.200		0.199	
	(0.000)		(0.000)	
Avg. percentile CSR ranking of firms in same state	0.157		0.163	
	(0.001)		(0.001)	
N	23,369	23,369	23,369	23,369
F-Statistic	63.18		58.35	
Wald Chi-squared		1,109.16		11,192.04
Anderson-Rubin overidentification test, p-value		0.0790		0.2190
Basmann overidentification, p-value		0.0815		0.2230
R-squared	8.75%	29.16%	9.07%	40.50%

REFERENCES

Amiraslani H., Lins K. V., Servaes H., Tamayo A. (2018). The bond market benefits of corporate social capital. Available at https://ssrn.com/abstract=2978794

Angrist, J.D., Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton: Princeton University Press.

Barber B. M., Lyon J. D. (1996). Detecting abnormal operating performance: The empirical power and specification of test statistics. *Journal of Financial Economics* 41(3): 359-399.

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

Certo S. T., Busenbark J. R., Kalm M., Lepine J. A. (2018). Divided we fall: How ratios undermine research in strategic management. *Organizational Research Methods*: https://doi.org/10.1177/1094428118773455.

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

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

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

Chung, K.H., Jo, H. (1996). The impact of security analysts' monitoring and marketing functions on the market value of firms. *Journal of Financial and Quantitative Analysis* 31: 493-512.

Durand, R., Paugam L., Stolowy, H., (2019). Do investors actually value sustainability indices? Replication, development, and new evidence on CSR visibility. *Strategic Management Journal*, forthcoming.

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

Fama E. F., French K. R. (1997). Industry costs of equity. *Journal of Financial Economics* 43(2): 153-193.

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

Flammer C. (2018). Corporate Green Bonds. https://ssrn.com/abstract=3125518.

Friedman M. (1970). The social responsibility of business is to increase its profits. New York Times Magazine, September, 13.

Godfrey PC. (2005). The relationship between corporate philanthropy and shareholder wealth: a risk management perspective. *Academy of Management Review* 30(4): 777–798.

Godfrey P. C., Merrill C. B., Hansen J. M. (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.

Gormley T. A., Matsa D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. *The Review of Financial Studies* 27(2): 617-661.

Hawn O., Chatterji A. K., Mitchell W. (2018). Do investors actually value sustainability? New evidence from investor reactions to the Dow Jones Sustainability Index (DJSI). *Strategic Management Journal* 39(4): 949-976.

Henderson, A. D., Raynor, M. E., Ahmed, M. (2012). How long must a firm be great to rule out chance? Benchmarking sustained superior performance without being fooled by randomness. *Strategic Management Journal* 33(4): 387-406.

Hong H., Liskovich I. (2015). Crime, punishment and the halo effect of corporate social responsibility, National Bureau of Economic Research: Available at http://www.nber.org/papers/w21215.

Husted B., Jamali D., Saffar W. (2016). Near and dear? The role of location in CSR engagement. *Strategic Management Journal* 37: 2050-2070.

Karpoff J., Schonlau R., Wehrly E. (2017). Do takeover defense indices measure takeover deterrence? *Review of Financial Studies* 30(7): 2359-2412.

Kempf A., Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management* 13(5). 908-922.

Kinder, Lydenberg, Domini. (2014). KLD - Social Index Methodology. *MSCI*.

Koh P. S., Qian C., Wang H. (2014). Firm litigation risk and the insurance value of corporate social performance. *Strategic Management Journal* 35(10): 1464-1482.

Lins K. V., Servaes H., Tamayo A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance* 72(4): 1785-1824.

Loughran T., Ritter J. R. (1997). The operating performance of firms conducting seasoned equity offerings. *The Journal of Finance* 52(5): 1823-1850.

Luo X., Wang H., Raithel S., Zheng Q. (2015). Corporate social performance, analyst stock recommendations, and firm future returns. *Strategic Management Journal* 36(1): 123-136.

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

Morck R., Shleifer A., Vishny R. W. (1988). Management ownership and market valuation: An empirical analysis. *Journal of Financial Economics* 20: 293-315.

Ramchander S., Schwebach R. G., Staking K. (2012). The informational relevance of corporate social responsibility: Evidence from DS400 Index reconstitutions. *Strategic Management Journal* 33(3): 303-314.

Servaes H., Tamayo A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science* 59(5): 1045-1061.

Shiu Y. M., Yang S. L. (2017). Does engagement in corporate social responsibility provide strategic insurance-like effects? *Strategic Management Journal* 38(2): 455-470.

SIF. (2018). Report on US sustainable, responsible and impact investing trends 2018. *Accessed via:https://www.ussif.org/files/Trends/Trends%202018%20executive%20summary%20FINAL.pdf (April 1, 2018)*.

Stock J., Yogo M. (2005). Testing for weak instruments in linear IV regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press: 80-108.

Verheyden, T., Eccles, R., Feiner, A. (2016) ESG for all? The impact of ESG screening on return, risk, and diversification. *Journal of Applied Corporate Finance* 28(2): 47-55.

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

Wang Z., Hsieh T. S., Sarkis J. (2018). CSR performance and the readability of CSR reports: Too good to be true? *Corporate Social Responsibility and Environmental Management* 25(1): 66-79.

World Bank (2017). Green Bond Impact Report 2017. *Investor Relations Department.*

World Bank (2018). Green Bond Impact Report 2018. *Investor Relations Department.*

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