Strategic Management Journal
Strat. Mgmt. J., **37**: 2378–2388 (2016)
Published online EarlyView 11 October 2016 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2579
Received 29 September 2014; *Final revision received 4 January 2016*

REVISITING THE CORPORATE SOCIAL PERFORMANCE-FINANCIAL PERFORMANCE LINK: A REPLICATION OF WADDOCK AND GRAVES

XIAOPING ZHAO[1]* and AUDREY J. MURRELL[2]
[1] *Department of Management and Organizations, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China*
[2] *College of Business Administration, University of Pittsburgh, Pittsburgh, Pennsylvania, U.S.A.*

Research summary: *In this study, we revisit the relationship between corporate social performance (CSP) and corporate financial performance (CFP) by conducting a replication of Waddock and Graves (1997). Using 1990 KLD ratings as the CSP measure, the original study reports a positive bidirectional relationship between CSP and CFP. However, our replication analyses with a larger sample over a longer time period indicate that the findings of the original study may not be generalizable to different samples. We argue that our replication casts doubt on the original study and can serve as a starting point to reconsider the CSP-CFP relationship. Based on the findings of our replication, we discuss the differences between the replication results and the original findings, and then suggest several approaches to revise and extend the original study.*

Managerial summary: *Advocates of corporate social performance (CSP) have long argued that "doing good leads to doing well." However, the evidence to support this argument is not strongly convincing, and managers hence doubt whether better CSP leads to improved corporate financial performance (CFP). In this article, we directly examine the relationship between CSP and CFP. Our article reports that CSP may not have a positive influence on CFP. Instead, our article shows the complexity of the relationship between CSP and CFP. Therefore, we cannot simply argue that doing good will necessarily lead to doing well.*

INTRODUCTION

Scholars have examined the relationship between corporate social performance (CSP) and corporate financial performance (CFP) for decades (for reviews, see Griffin and Mahon, 1997, Margolis and Walsh, 2003, and Orlitzky, Schmidt, and Rynes, 2003). According to these studies, CSP has strategic implications for firms because better CSP is likely to lead to improved financial performance. Drawing on previous studies, scholars generally believe that CSP is positively related to CFP, although the relationship is modest (Wood, 2010). However, recent studies have begun to question this generally held belief because prior studies on CSP-CFP relationships suffer from several conceptual and methodological problems (e.g., Bridoux and Stoelhorst, 2014; Garcia-Castro, Ariño, and Canela, 2010; Harrison and Bosse, 2013). In this study, we revisit the relationship between CSP and CFP by replicating Waddock and Graves's (1997) research published in *Strategic Management Journal*.

Drawing from stakeholder theory (Freeman, 1984) and slack resource arguments (Nohria and Gulati, 1996), Waddock and Graves (1997: 307) propose that (1) "better financial performance results in improved CSP, *ceteris paribus*," and

Keywords: corporate social performance (CSP); corporate financial performance (CFP); CSP-CFP relationship; replication; stakeholder theory
*Correspondence to: Xiaoping Zhao, Department of Management and Organizations, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China. E-mail: zhaoxiaoping@sjtu.edu.cn

that (2) "improved CSP leads to better financial performance, *ceteris paribus*." Using 1990 KLD ratings as the CSP measure and three accounting performance measures (ROA, ROE, and ROS), Waddock and Graves (1997) found empirical support for their arguments. We replicate Waddock and Graves's (1997) research for two reasons. First, their research is the most cited study addressing the CSP-CFP relationship and has been cited more than 3,200 times between when it was published and March 2015, according to Google Scholar, and approximately 1,600 times since 2011. More important, some articles continue to build their conceptual and methodological bases on Waddock and Graves (1997), including Cheng, Ioannou, and Serafeim (2014), Chun *et al.* (2013), Koh, Qian, and Wang (2014), Marquis and Qian (2014), Wang and Bansal (2012), and Wang and Qian (2011). Second, Peloza's (2009) review of CSP measures revealed numerous studies that follow the Waddock and Graves (1997) approach, and used KLD ratings to measure CSP. Recent studies continue to use KLD ratings to measure CSP (e.g., Jayachandran, Kalaignanam, and Eilert, 2013; Koh *et al.*, 2014; Luo *et al.*, 2015; Ramchander, Schwebach, and Staking, 2012), despite the ongoing debate on whether KLD ratings are good measures of CSP (Chatterji, Levine, and Toffel, 2009; Mattingly and Berman, 2006).

To replicate Waddock and Graves's (1997) study, we use the same measures and similar analytical procedures as they do, but expand the time frame and use a substantially larger sample (KLD ratings for the period 1991–2013). The purpose of this replication is not to reproduce the original results reported by Waddock and Graves, and thus, we do not re-analyze the original study's data set. Instead, our major objective for this replication is to employ the same specifications and tests to examine the generalizability of the original study to different settings, time periods, and places. Analyzing more than 27,000 firm-year observations from 1991 to 2013, our replication analyses examine the relationship between CSP and accounting performance measured as ROA, ROE, and ROS, like that reported by Waddock and Graves (1997). In addition, our replication analyses include financial market performance, measured by Tobin's q, market-to-book ratio (MTB), and market value added (MVA), and its relationship with CSP. Our findings with a larger sample over a longer time period differ somewhat from the original findings. Specifically, our replication analyses confirm the original finding that prior financial performance is related to subsequent CSP, but cast doubts on another original finding that prior CSP is positively associated with subsequent financial performance. More important, the effect sizes of the CSP-CFP associations yielded by our replication analyses are substantially smaller than those reported by the original study. To explain the differences between the original findings and our replication results, we discuss the substantial differences in the original sample and the samples of our replication analyses. The empirical evidence offered by our replication analyses suggests that the relationship between CSP and CFP merits future study. Thus, we propose a few approaches for future research intended to capture the complex nature of the CSP-CFP relationship.

REPLICATION METHODS

Because our primary purpose is to replicate Waddock and Graves's (1997) study using data from a longer period to determine whether a different sample generates different results with the same specifications and tests, we did not re-analyze the exact data used by the original study. However, we employed the same measures and similar estimation method used by the original study. In essence, regression analyses including firm and year fixed-effects were used to examine the relationship between CSP and CFP. However, due to the presence of outliers or extreme values of the CFP measures (i.e., ROA, ROE, ROS, Tobin's q, MTB, and MVA), we reconducted the fixed-effects regression analyses when these extreme values were removed from the replication samples. We suspected that the empirical results might be biased by the presence of extreme values of these variables. After checking the distributions and histograms of these variables, we used the 5th and 95th percentile of CFP as cut-offs for extreme values.

Sample and measures

Sample

Waddock and Graves (1997) only used KLD data from 1990, which contained 469 observations. Our initial sample consisted of all the firms in the 1991–2013 KLD ratings. Then, we combined this

DOI: 10.1002/smj

data set with Compustat data, which was the source for the CFP and control variables. To examine the influence of CFP on CSP, we combined 1991–2013 KLD data with 1990–2012 Compustat data. In contrast, to examine the influence of CSP on CFP, we combined 1991–2013 KLD data with 1992–2014 Compustat data. As a result, we generated two replication samples, both of which were based on 1991–2013 KLD data. However, due to the missing data on CFP measures and control variables in Compustat, the two replication samples were not identical, although they had substantial overlap in firm-year observations in terms of firm identity and year. Specifically, replication sample 1, with 29,352 firm-year observations, was used to examine the influence of CFP on CSP, whereas replication sample 2, with 27,025 firm-year observations, was used to examine the influence of CSP on CFP. The major reason that replication sample 1 was larger than replication sample 2 was that when the replication analyses were conducted, considerable 2014 CFP data were missing from Compustat. In particular, replication sample 1 contained 3,850 unique firm-year observations that were not included in replication sample 2; similarly, replication sample 2 contained 1,523 unique firm-year observations that were not included in replication sample 1. In this sense, the two replication samples contained 25,502 firm-year observations in common. To test the robustness of our replication results, we also conducted additional analyses using the sample with 25,502 firm-year observations. We found that the results using this sample did not differ substantially from replication results using replication sample 1 and replication sample 2.

CSP

Following the Waddock and Graves (1997) approach, CSP was measured as a weighted average score of eight KLD social rating dimensions. In the original study, CSP was measured as a weighted average score of eight KLD social rating dimensions (i.e., employee relations, product, community relations, environment, treatment of women and minorities, nuclear power, military contracts, and South Africa). The weights of these eight dimensions were generated by the simple multi-attribute rating technique (SMART; Von Winterfeldt and Edwards, 1986). The authors of the original study asked three CSP experts to evaluate the importance of each CSP dimension and to assign them weights. Using this weighting scale, the authors of the original study calculated a weighted average CSP score for each firm in their sample. We adopted the same weights reported by Waddock and Graves (1997: 310) to calculate the weighted average CSP score.

CFP

Following the original study, we used three accounting performance measures: ROA, ROE, and ROS. However, we added three financial market performance measures (i.e., Tobin's q, MTB, and MVA) to determine whether CSP was correlated with these financial market performance indicators, which were not readily available for firms in the original sample. For the accounting performance measures, ROA was calculated as net income over total assets; ROE was calculated as net income over total equity; and ROS was calculated as net income over total sales. The scales of the three measures in our replication analyses were the same as the original study. For the market performance measures, Tobin's q was calculated by the following formula: (equity market value + debt market value)/(equity book value + debt book value); MTB was calculated as equity market value over equity book value; and MVA was calculated as equity market value—book value of equity and debt.

Control variables

Following the original study's methodology, control variables included the ratio of debt to total assets, firm size (measured as total sales, total assets, and number of employees), and a set of year dummy variables. Consistent with the original study, there was a one-year lag between the dependent variable and independent/control variables.

REPLICATION RESULTS

Table 1 presents the descriptive statistics. Panel 1 of Table 1 reports the descriptive statistics of the original sample adopted from Waddock and Graves (1997: 309, Table 2). Panels 2 and 3 show the descriptive statistics of the replication samples. Specifically, Panel 2 reports the descriptive statistics of the sample we used to examine the influence of CFP on CSP; in contrast, Panel 3 reports the

DOI: 10.1002/smj

descriptive statistics of the sample we used to examine the influence of CSP on CFP.

Table 2 presents the correlation matrix for variables in our replication samples. Panel 1 of Table 2 reports the correlations when CSP is the dependent variable and CFP is the independent variable. Panel 2 of Table 2 reports the correlations when CFP is the dependent variable and CSP is the independent variable. In both samples, the three measures of firm size (i.e., total sales, total assets, and number of employees) are highly correlated. As in the original study, they are entered into the regression models separately.

The associations between prior CFP on subsequent CSP

Tables 3 and 4 report the empirical results of our replication analyses.[1] In particular, Table 3 reports the empirical results when CSP is the dependent variable and CFP is the independent variable. Panel 1 reports the regression coefficients and the ranges of *p*-values adopted from the original study (Waddock and Graves, 1997: 312, Table 5), and Panel 2 reports the regression coefficients, robust standard errors, and exact *p*-values of our replication analyses. Here, the measures of CFP include ROA, ROE, ROS, Tobin's q, MTB, and MVA. Following the analytical procedures of the original study (Waddock and Graves, 1997), these variables are entered into the replication analyses separately.

According to Panel 1 of Table 3, the original finding is that ROA is positively related with CSP. Panel 2 shows that our replication analyses yield positive but noticeable smaller associations between ROA and CSP when the extreme values of ROA are removed. For example, as firm size is measured by total sales, the effect size reported by the original study is 1.189 ($p < 0.001$), while the effect size reported by our replication analyses is 0.242 (S.E. = 0.043, $p = 0.000$). Similarly, when firm size is measured by total assets or number of employees, the effect sizes of the associations between ROA and CSP reported by our replication analyses are extremely smaller than those reported by the original study.

With respect to the associations between CSP and ROE, both the original study and our replication

Table 1. Descriptive statistics of original sample and replication samples

Variable	Panel 1: original sample (Waddock and Graves, 1997)		Panel 2: replication sample 1 (DV = CSP, IV = CFP) CSP: 1991–2013 CFP and control variables: 1990–2012				Panel 3: replication sample 2 (DV = CFP, IV = CSP) CFP: 1992–2014 CSP and control variables: 1991–2013			
	Mean	Std. dev.	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
CSP	−0.034	0.356	−0.015	0.335	−1.347	2.462	−0.023	0.333	−1.347	2.462
ROA	0.055	0.058	0.020	0.205	−8.681	9.738	−0.004	3.189	−523.200	2.408
ROE	0.139	0.283	0.003	6.536	−790.608	141.742	0.033	4.995	−622.383	142.787
ROS	0.059	0.073	−2.903	198.017	−29319.000	609.074	−3.138	206.179	−29319.000	110.597
Tobin's q	–	–	2.261	16.358	0.345	1,979.617	2.036	8.259	0.330	1,219.992
MTB	–	–	3.554	64.666	−4,027.241	7,426.010	2.874	36.261	−4,027.241	1,575.700
MVA	–	–	−4,310.647	62,375.630	−2,232,744.000	559,320.000	−4,907.623	68,759.410	−2,232,744.000	559,320.000
Debt to total assets	0.203	0.174	0.556	0.272	0.006	6.189	0.556	0.267	0.005	4.072
Total sales	6,416.370	11,133.740	4,676.553	15,853.050	0.000	444,948	4,950.219	16,867.750	0.000	467.231
	M$	M$	M$	M$	M$	M$	M$	M$	M$	M$
Total assets	11,444.750	23,598.730	10,648.690	68,986.230	0.396	2,400,000	11,280.760	71,584.410	0.486	2,300,000
	M$	M$	M$	M$	M$	M$	M$	M$	M$	M$
Employees	39,646 thous	64,358 thous	16,900 thous	58,512 thous	0.001 thous	2,200 thous	17,539 thous	61,475 thous	0.001 thous	2,200 thous
Total observations	469		29,352				27,025			

[1] The complete results with all of the estimates, including control variables, can be found in Appendix S1.

DOI: 10.1002/smj

Table 2. Correlations of variables of replication samples

Replication sample 1: CSP (DV): 1991–2013; CFP (IV) and control variables: 1990–2012											
Variable		1	2	3	4	5	6	7	8	9	10
1.	CSP										
2.	ROA	0.066									
3.	ROE	0.000	0.068								
4.	ROS	−0.002	0.055	0.003							
5.	Tobin's q	0.013	−0.022	0.001	−0.001						
6.	MTB	0.004	0.003	−0.398	0.000	0.329					
7.	MVA	−0.068	0.022	0.001	−0.001	0.014	0.007				
8.	Debt to total assets	0.009	−0.119	−0.011	−0.003	−0.011	−0.008	−0.129			
9.	Total sales	0.132	0.045	0.006	0.004	−0.008	−0.002	−0.282	0.110		
10.	Total assets	0.135	0.004	0.002	0.002	−0.008	−0.003	−0.948	0.127	0.472	
11.	Employees	0.113	0.046	0.006	0.004	−0.006	−0.001	−0.141	0.084	0.689	0.273
Replication sample 2: CFP (DV): 1992–2014; CSP (IV) and control variables: 1991–2013											
Variable		1	2	3	4	5	6	7	8	9	10
1.	CSP										
2.	ROA	0.004									
3.	ROE	−0.003	0.003								
4.	ROS	0.000	0.010	0.004							
5.	Tobin's q	0.018	−0.897	0.000	−0.008						
6.	MTB	0.010	0.000	0.184	0.000	0.057					
7.	MVA	−0.066	0.001	0.001	−0.001	0.023	0.010				
8.	Debt to total assets	0.007	−0.055	−0.013	−0.003	0.022	−0.022	−0.134			
9.	Total sales	0.110	0.005	0.009	0.005	−0.010	0.003	−0.274	0.110		
10.	Total assets	0.130	0.001	0.003	0.002	−0.013	−0.003	−0.941	0.131	0.468	
11.	Employees	0.102	0.005	0.008	0.004	−0.005	0.005	−0.139	0.084	0.693	0.271

Replication sample 1: N = 29,352; $p < 0.05$ if $|r| > 0.012$.
Replication sample 2: N = 27,025; $p < 0.05$ if $|r| > 0.013$.

analyses report positive associations, but the effect sizes in our replication analyses are smaller. For instances, as firm size is measured by total sales, the effect size reported by the original study is 0.110 ($p < 0.100$), while the effect size reported by our replication analyses is 0.060 (S.E. = 0.019, $p = 0.002$). In addition, when firm size is measured by total assets or number of employees, the effect sizes of the associations between ROE and CSP reported by our replication analyses are smaller than those reported by the original study.

A similar pattern has also been founded regarding the association between ROS and CSP. Specifically, as firm size is measured by total sales, the effect size reported by the original study is 0.597 ($p < 0.050$), while the effect size reported by our replication analyses is 0.150 (S.E. = 0.023, $p = 0.000$). In addition, when firm size is measured by total assets or number of employees, the effect sizes of the associations between ROS and CSP reported by our replication analyses are smaller than those reported by the original study.

Panel 2 of Table 3 also reports the results with respect to the associations between CSP and each of three financial market performance measures (i.e., Tobin's q, MTB, and MVA). In general, our replication results show that the associations between financial market performance and CSP are generally positive after removing extreme values of financial market performance measures. However, we note that the effect sizes are very small. For example, when firm size is measured as total sales, the effect size of the association between Tobin's q and CSP is 0.010 (S.E. = 0.005, $p = 0.035$); the effect size of the relationship between MTB and CSP is 0.005 (S.E. = 0.002, $p = 0.027$); and the effect size of the relationship between MVA and CSP is 5.43E-06 (S.E. = 2.05E-06, $p = 0.008$). Moreover, when firm size is measured by total assets or number of employees, the effect sizes of the associations between CSP and financial market performance measures remain very small.

To summarize, after removing extreme values of performance measures, our replication results

DOI: 10.1002/smj

Table 3. The influence of CFP on CSP

	Firm size measure: total sales			Firm size measure: total assets			Firm size measure: number of employees		
	Coefficients	Robust S.E.	*p*-Value	Coefficients	Robust S.E.	*p*-Value	Coefficients	Robust S.E.	*p*-Value
Panel 1: OLS coefficients adopted from Table 5 in Waddock and Graves (1997: 312)									
IV = ROA, DV = CSP	1.189	–	$p < 0.001$	1.206	–	$p < 0.001$	1.225	–	$p < 0.001$
IV = ROE, DV = CSP	0.110	–	$p < 0.10$	0.108	–	$p < 0.10$	0.109	–	$p < 0.10$
IV = ROS, DV = CSP	0.597	–	$p < 0.05$	0.602	–	$p < 0.05$	0.614	–	$p < 0.05$
Panel 2: replication using 1991–2013 KLD data									
IV = ROA, DV = CSP									
1. Fixed-effects	0.018	0.008	0.031	0.019	0.008	0.026	0.017	0.008	0.036
2. Fixed-effects, no extreme values	0.242	0.043	0.000	0.249	0.042	0.000	0.241	0.042	0.000
IV = ROE, DV = CSP									
1. Fixed-effects	−8.51E-05	1.43E-04	0.551	−8.28E-05	1.42E-04	0.559	−8.80E-05	1.41E-04	0.532
2. Fixed-effects, no extreme values	0.060	0.019	0.002	0.066	0.019	0.000	0.061	0.019	0.001
IV = ROS, DV = CSP									
1. Fixed-effects	−6.01E-06	8.62E-07	0.000	−5.99E-06	8.62E-07	0.000	−5.91E-06	8.69E-07	0.000
2. Fixed-effects, no extreme values	0.150	0.023	0.000	0.153	0.023	0.000	0.151	0.023	0.000
IV = Tobin's q, DV = CSP									
1. Fixed-effects	2.11E-04	8.90E-05	0.018	2.10E-04	8.89E-05	0.019	2.12E-04	9.22E-05	0.021
2. Fixed-effects, no extreme values	0.010	0.005	0.035	0.010	0.005	0.044	0.008	0.005	0.104
IV = MTB, DV = CSP									
1. Fixed-effects	1.56E-05	2.07E-05	0.451	1.56E-05	2.06E-05	0.451	1.63E-05	2.11E-05	0.440
2. Fixed-effects, no extreme values	0.005	0.002	0.027	0.005	0.002	0.024	0.004	0.002	0.073
IV = MVA, DV = CSP									
1. Fixed-effects	−3.36E-07	1.70E-07	0.048	5.85E-07	3.64E-07	0.108	−4.46E-07	1.59E-07	0.005
2. Fixed-effects, no extreme values	5.43E-06	2.05E-06	0.008	7.49E-06	2.14E-06	0.000	4.20E-06	2.09E-06	0.044

1. We used 5th and 95th percentile of CFP as cutoffs for extreme values. After dropping extreme values of CFP, the sample size was 26,416, compared with 29,352 of the full sample.
2. We also used 1st and 99th percentile of CFP as cutoffs for extreme values. After dropping extreme values of CFP, the sample size was 24,991, compared with 25,502 of the full sample. The empirical results between the two ways to address extreme values differ qualitatively for ROS, Tobin's q, MTB, and MVA. By checking the distributions and histograms of these variables, we found that using 1st and 99th percentile could not deal with many extreme values, and hence, using 5th and 95th percentile was more appropriate.

DOI: 10.1002/smj

Table 4. The influence of CSP on CFP

	Firm size measure: total sales			Firm size measure: total assets			Firm size measure: number of employees		
	Coeficients	Robust S.E.	*p*-value	Coeficients	Robust S.E.	*p*-value	Coeficients	Robust S.E.	*p*-value
Panel 1: OLS coefficients adopted from Table 6 in Waddock and Graves (1997: 313)									
DV = ROA, IV = CSP	0.024	–	$p < 0.01$	0.024	–	$p < 0.001$	0.024	–	$p < 0.01$
DV = ROE, IV = CSP	0.081	–	$p > 0.10$	0.081	–	$p > 0.10$	0.081	–	$p > 0.10$
DV = ROS, IV = CSP	0.021	–	$p < 0.05$	0.021	–	$p < 0.05$	0.022	–	$p < 0.05$
Panel 2: replication using 1991 – 2013 KLD data									
IV = CSP, DV = ROA									
1. Fixed-effects	−0.003	0.004	0.425	−0.003	0.004	0.462	−0.003	0.004	0.370
2. Fixed-effects, no extreme values	0.004	0.002	0.013	0.005	0.002	0.010	0.004	0.002	0.018
IV = CSP, DV = ROE									
1. Fixed-effects	−0.054	0.116	0.643	−0.053	0.117	0.652	−0.055	0.116	0.634
2. Fixed-effects, no extreme values	0.001	0.004	0.767	0.002	0.004	0.572	0.001	0.004	0.791
IV = CSP, DV = ROS									
1. Fixed-effects	−0.059	0.662	0.929	−0.053	0.664	0.936	−0.016	0.643	0.980
2. Fixed-effects, no extreme values	0.003	0.003	0.201	0.004	0.003	0.159	0.003	0.003	0.225
IV = CSP, DV = Tobin's q									
1. Fixed-effects	−0.079	0.066	0.229	−0.083	0.067	0.218	−0.098	0.067	0.142
2. Fixed-effects, no extreme values	0.002	0.025	0.949	0.001	0.026	0.973	−0.010	0.025	0.681
IV = CSP, DV = MTB									
1. Fixed-effects	0.023	0.557	0.968	0.033	0.562	0.953	0.001	0.556	0.999
2. Fixed-effects, no extreme values	0.022	0.058	0.705	0.030	0.059	0.616	−0.001	0.059	0.982
IV = CSP, DV = MVA									
1. Fixed-effects	−8,065.098	6,018.326	0.180	265.930	897.453	0.767	−10,821.130	6,998.218	0.122
2. Fixed-effects, no extreme values	173.041	125.394	0.168	211.292	121.334	0.082	103.846	127.223	0.414

1. We used 5th and 95th percentile of CFP as cutoffs for extreme values. After dropping extreme values of CFP, the sample size was 24,332, compared with 27,025 of the full sample.
2. We also used 1st and 99th percentile of CFP as cutoffs for extreme values. After dropping extreme values of CFP, the sample size was 26,483, compared with 27,065 of the full sample. The empirical results between the two ways to address extreme values do not differ qualitatively. In addition, by checking the distributions and histograms of such variables as ROS, Tobin's q, MTB, and MVA, we found that using 1st and 99th percentile could not deal with many extreme values, and hence, using 5th and 95th percentile was more appropriate.

DOI: 10.1002/smj

confirm the original findings that prior accounting performance measured as ROA, ROE, and ROS) is positively associated with subsequent CSP (Waddock and Graves, 1997). However, we note that the effect sizes reported by our replication analyses are extremely smaller than those reported by the original study. Furthermore, our replication analyses indicate that prior financial market performance measured as Tobin's q, MTB, and MVA could be positively related with subsequent CSP; however, the effect sizes reported by our replication analyses are small.

The associations between prior CSP and subsequent CFP

Table 4 reports the empirical results when CFP is the dependent variable and CSP is the independent variable. Panel 1 reports the regression coefficients and the ranges of *p*-values obtained from the original study (Waddock and Graves, 1997: 313, Table 6), and Panel 2 reports the regression coefficients, robust standard errors, and exact *p*-values of our replication analyses. Here, the measures of CFP include ROA, ROE, ROS, Tobin's q, MTB, and MVA. Following the analytical procedures of the original study (Waddock and Graves, 1997), these variables are entered into the replication analyses separately.

According to Panel 1 of Table 4, the original finding is that CSP is positively related with ROA. Panel 2 shows that our replication analyses yield positive but smaller associations between CSP and ROA when the extreme values of ROA are removed. For example, as firm size is measured by total sales, the effect size reported by the original study is 0.024 ($p < 0.01$), while the effect size reported by our replication analyses is 0.004 (S.E. = 0.002, $p = 0.013$). Similarly, when firm size is measured by total assets or number of employees, the effect sizes reported by our replication analyses are positive but smaller than those reported by the original study.

With respect to the associations between CSP and ROE, both the original study and our replication analyses demonstrate associations that do not significantly differ from zero. But the effect sizes yielded by our replication analyses are smaller than those of the original study. For instances, as firm size is measured by total sales, the effect size reported by the original study is 0.081 ($p > 0.10$), while the effect size reported by our replication analyses is 0.001 (S.E. = 0.004, $p = 0.767$). In addition, when firm size is measured by total assets or number of employees, the effect sizes reported by our replication analyses are still substantially smaller than those reported by the original study.

The original study reports a positive relationship between CSP and ROS, whereas our replication analyses show associations that do not significantly differ from zero. Specifically, as firm size is measured by total sales, the effect size reported by the original study is 0.021 ($p < 0.050$), while the effect size reported by our replication analyses is 0.003 (S.E. = 0.003, $p = 0.201$). In addition, when firm size is measured by total assets or number of employees, the effect sizes reported by our replication analyses also does not significantly differ from zero and are extremely smaller than those reported by the original study.

Panel 2 of Table 4 also reports the results with respect to the associations between CSP and each of three financial market performance measures (i.e., Tobin's q, MTB, and MVA). In general, our replication results show that the associations between CSP and financial market performance are generally positive, but do not significantly differ from zero after removing extreme values of financial market performance measures. For example, when firm size is measured as total sales, the effect size of the association between CSP and Tobin's q is 0.002 (S.E. = 0.025, $p = 0.949$); the effect size of the relationship between CSP and MTB is 0.022 (S.E. = 0.058, $p = 0.705$); and the effect size of the relationship between CSP and MVA is 173.041 (S.E. = 125.394, $p = 0.168$). Moreover, when firm size is measured by total assets or number of employees, the associations between CSP and financial market performance measures still do not significantly differ from zero.

To summarize, after removing extreme values of the accounting performance measures, our replication analyses yield positive associations between ROA and CSP. However, the effect sizes of such positive associations are substantially smaller than the effect sizes reported by the original study. In addition, with respect to the associations between CSP and the two other accounting performance measures (i.e., ROE and ROS), our replication analyses report associations that do not significantly differ from zero and the effect sizes yielded by our replication analyses are also smaller than those reported by the original study. Moreover, our replication results show that the associations between CSP and financial market performance do

DOI: 10.1002/smj

not significantly differ from zero. In essence, our replication results could not support the original argument that prior CSP is positively associated with subsequent CFP (Waddock and Graves, 1997).

DISCUSSION AND CONCLUSION

In this article, we revisit the relationship between CSP and CFP by replicating Waddock and Graves's (1997) study. The core theme of their research (1997: 307) is that there is a virtuous circle in which "*better financial performance results in improved CSP*" and "*improved CSP leads to better financial performance*." Our replication results partially support Waddock and Graves's (1997) arguments. Specifically, our replication results provide evidence that better accounting performance is associated with improved CSP; however, the effect size of this association yielded by our replication analyses is noticeably smaller than that reported by Waddock and Graves (1997). Furthermore, our replication results cast doubts on the argument that improved CSP is related to better accounting performance. In sum, our replication results show that the positive association between prior CSP on subsequent CFP argued by Waddock and Graves (1997) is not generalizable to a different sample that includes more firms over a longer time period, whereas the positive association between prior CFP and subsequent CSP reported by the original study could be generalizable. More important, the effect sizes yielded by our replication analyses are substantially smaller than those reported by the original study.

Although previous reviews argue that sample selection may cause problems with exploring CSP-CFP relationships (e.g., Margolis and Walsh, 2003), to our knowledge, our replication is the first study to test this notion and demonstrate that the positive association between prior CSP and subsequent CFP found in one sample may not be generalizable to other samples. In this sense, one explanation is that the original findings obtained from one sample may not carry over to another (Hamermesh, 2007). Thus, differences between the replication findings and the original results could be caused by differences between the original sample (KLD 1990) and the replication samples (KLD 1991–2013). For example, the samples used in our replication differ substantially from that of the original study in several aspects (see Table 1). First, the sample used in the original study only covers 469 S&P 500 firms, while the samples employed in our replication cover more than 27,000 firm-year observations. This is because beginning in 2001, KLD substantially expanded the coverage of rated firms, from only firms covered by the *S&P 500 Index* and the *Domini 400 Social ^{SM} Index* to other firms that are listed in *1,000 Largest US Companies*, *Large Cap Social ^{SM} Index*, *2,000 Small Cap US Companies*, and *Broad Market Social ^{SM} Index*. In general, KLD ratings covered approximately 650 firms between 1991 and 2000 compared with more than 3,000 firms after 2003. A second difference is that firms in the replication samples have lower profitability (i.e., ROA, ROE, and ROS), relatively smaller size (i.e., total sales, total assets, and employees), and higher debt-to-assets ratios than firms in the original sample (see Table 1). Because the replication samples differ from the original sample in certain dimensions that might impact the relationship between CSP and CFP, one might argue that these differences would impact the relationship between CSP and CFP observed in our replication analyses (e.g., Wang and Qian, 2011).

Theoretical implications for future research

In general, we suggest that two specific results of our replication analyses merit discussion. First, our replication analyses report a mixed CSP-CFP relationship. Specifically, we find that prior CFP is positively related to subsequent CSP, but prior CSP is not associated with subsequent CFP. Second, the effect sizes of CSP-CFP relationship yielded by our replication analyses are noticeably smaller than those reported by the original study. Although the discovery of a mixed CSP-CFP relationship is not a unique finding of our replication analyses (e.g., Margolis and Walsh, 2003; Orlitzky *et al.*, 2003) and it is beyond the scope of a replication study to comprehensively explore the many reasons that may explain the mixed CSP-CFP relationship, we suggest that the two replication results noted above have important implications for future study. That is, future study may examine the conditions under which prior CSP is associated with subsequent CFP and also pay attention to the effect sizes of CSP-CFP relationship.

To do so, future study may directly examine the mediating effects of stakeholder responses between CSP and CFP. Waddock and Graves (1997) theoretically argue that better CSP improves subsequent CFP because better CSP establishes

DOI: 10.1002/smj

good stakeholder relationships, which, in turn, enhance employee engagement and satisfaction, reduce governmental regulations, and result in greater customer purchasing intentions and behaviors; consequently, these positive stakeholder responses to better CSP lead to improved CFP. However, our replication results show that prior CSP is not positively associated with subsequent CFP, implying that stakeholders may not necessarily positively respond to better CSP (Bridoux and Stoelhorst, 2014; Harrison and Bosse, 2013). In fact, a few recent studies have shown that better CSP may not enhance employee engagement and satisfaction when employees are treated unfairly or are low in moral identity (Rupp *et al.*, 2013) or when the support for CSP from top management is low (Mallory and Rupp, 2015). In addition, better CSP may not enhance customer purchasing intentions and behaviors when customers perceive that a firm does not seek to improve CSP to enhance stakeholders' welfare but to generate more profits or when the degree of logical fit between the firm and its CSP is low (for a review, see Peloza and Zhang, 2011). In addition to exploring the variations of stakeholder responses to CSP, future study may also examine the variability in the magnitude of stakeholder responses. As discussed, the effect sizes of CSP-CFP relationship yielded by our replication analyses are substantially smaller than those reported by Waddock and Graves (1997). We suspect that the magnitude of stakeholder responses to CSP matters to the effect size of CSP-CFP relationship because stronger stakeholder responses to CSP could have a stronger influence on CFP. Thus, future study may explore this possibility.

ACKNOWLEDGEMENTS

This research was supported by funding provided by the University of Pittsburgh, David Berg Center for Ethics and Leadership, and through a grant from the Bank of New York Mellon Foundation.

REFERENCES

Bridoux F, Stoelhorst J. 2014. Microfoundations for stakeholder theory: managing stakeholders with heterogeneous motives. *Strategic Management Journal* **35**(1): 107–125.

Chatterji AK, Levine DI, Toffel MW. 2009. How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy* **18**(1): 125–169.

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

Chun JS, Shin Y, Choi JN, Kim MS. 2013. How does corporate ethics contribute to firm financial performance? The mediating role of collective organizational commitment and organizational citizenship behavior. *Journal of Management* **39**(4): 853–877.

Freeman RE. 1984. *Strategic Management: A Stakeholder Approach*. Pitman: Boston, MA.

Garcia-Castro R, Ariño MA, Canela MA. 2010. Does social performance really lead to financial performance? Accounting for endogeneity. *Journal of Business Ethics* **92**(1): 107–126.

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

Hamermesh DS. 2007. Viewpoint: replication in economics. *Canadian Journal of Economics* **40**(3): 715–733.

Harrison JS, Bosse DA. 2013. How much is too much? The limits to generous treatment of stakeholders. *Business Horizons* **56**(3): 313–322.

Jayachandran S, Kalaignanam K, Eilert M. 2013. Product and environmental social performance: varying effect on firm performance. *Strategic Management Journal* **34**(10): 1255–1264.

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

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.

Mallory DB, Rupp DE. 2015. "Good" leadership: using corporate social responsibility to enhance leader-member exchange. In *The Oxford Handbook of Leader-Member Exchange*, Bauer IT, Erdogan B (eds). Oxford University Press: Oxford, UK; 335–350.

Margolis JD, Walsh JP. 2003. Misery loves companies: rethinking social initiatives by business. *Administrative Science Quarterly* **48**(2): 268–305.

Marquis C, Qian C. 2014. Corporate social responsibility reporting in China: symbol or substance? *Organization Science* **25**(1): 127–148.

Mattingly JE, Berman SL. 2006. Measurement of corporate social action discovering taxonomy in the Kinder Lydenburg Domini ratings data. *Business and Society* **45**(1): 20–46.

Nohria N, Gulati R. 1996. Is slack good or bad for innovation? *Academy of Management Journal* **39**(5): 1245–1264.

Orlitzky M, Schmidt FL, Rynes SL. 2003. Corporate social and financial performance: a meta-analysis. *Organization Studies* **24**(3): 403–441.

Peloza J. 2009. The challenge of measuring financial impacts from investments in corporate social performance. *Journal of Management* **35**(6): 1518–1541.

DOI: 10.1002/smj

Peloza J, Zhang J. 2011. How can corporate social responsibility activities create value for stakeholders? A systematic review. *Journal of the Academy of Marketing Science* **39**(1): 117–135.

Ramchander S, Schwebach RG, Staking K. 2012. The informational relevance of corporate social responsibility: evidence from DS400 index reconstitutions. *Strategic Management Journal* **33**(3): 303–314.

Rupp DE, Shao R, Thornton MA, Skarlicki DP. 2013. Applicants' and employees' reactions to corporate social responsibility: the moderating effects of first-party justice perceptions and moral identity. *Personnel Psychology* **66**(4): 895–933.

Von Winterfeldt D, Edwards W. 1986. *Decision Analysis and Behavioral Research*. Cambridge University Press: Cambridge, UK.

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

Wang T, Bansal P. 2012. Social responsibility in new ventures: profiting from a long-term orientation. *Strategic Management Journal* **33**(10): 1135–1153.

Wang H, Qian C. 2011. Corporate philanthropy and corporate financial performance: the roles of stakeholder response and political access. *Academy of Management Journal* **54**(6): 1159–1181.

Wood DJ. 2010. Measuring corporate social performance: a review. *International Journal of Management Reviews* **12**(1): 50–84.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix S1. Complete Replication Results.

DOI: 10.1002/smj