



WHY IS THE FINANCIAL PERFORMANCE-ENVIRONMENTAL PERFORMANCE RELATIONSHIP DIFFICULT TO MEASURE?

DOI: 10.17261/Pressacademia.2018.945

RJBM-V.5-ISS.3-2018(4)-p.212-221

Kenneth J. Hatten¹, James P. Keeler², William L. James³, Kyungho Kim⁴

¹Boston University, Questrom School of Business, 595 Commonwealth Avenue, Boston, MA 02215, USA.

kjhatten@bu.edu, ORCID: 0000-0003-3077-0965

²Kenyon College, Department of Economics, Gambier, OH 43022, USA

keeler@kenyon.edu, ORCID: 0000-0002-6804-6756

³Hofstra University, Zarb School of Business, Hempstead, NY 11549, USA.

william.james@hofstra.edu, ORCID: 0000-0002-5718-7879

⁴Ajou University, Korea.

kyunghokim@ajou.ac.kr, ORCID: 0000-0002-1477-1266

Date Received: March 7, 2018

Date Accepted: September 17, 2018

To cite this document

Hatten, K., Keeler, J. P., James, W. L., Kim, K. (2018). Why is the financial performance-environmental performance relationship difficult to measure?. Research Journal of Business and Management (RJBM), 5(3), p.212-221.

Permanent link to this document: <http://doi.org/10.17261/Pressacademia.2018.945>

Copyright: Published by PressAcademia and limited licenced re-use rights only.

ABSTRACT

Purpose- There appears to be no consensus as to the nature (positive or negative) or strength of the relationship between financial performance (FP) and environmental performance (EP). The literature seems to lean towards a positive relationship. This paper offers an explanation of the mixed empirical results.

Methodology- Using publicly available data a regression model with control variables was developed. Serial correlation was present and the model was adjusted to account for this. Quantile regression allowed a more direct test of our hypotheses by allowing a direct estimate, with less bias, of how the EP-FP relation varies by level of EP.

Findings- This paper explores this relationship within three U.S. industries. The relationship as modeled here is serially correlated and differs across industries and time. Moreover, within three different single industries, quantile regression shows that the relationship differs for high and low polluters.

Conclusion- This research indicates that the strength of the relationship between financial performance and environmental performance is weak and varies considerably depending on level of environmental performance and industry. We conclude that the mixed results in the literature of the EP – FP relation are due to failures to address serial correlation bias and heterogeneity across industries and degree of environmental performance.

Keywords: Financial performance, environmental performance, conflicting results, quantile regression, research methods.

JEL Codes: L25, M21, Q58

1. INTRODUCTION

The focus here is trying to eliminate uncertainty about the financial and environmental performance (FP-EP) relationship and its sign. Fundamentally, we question whether the models used in a great amount of published research were properly specified. Consistent with this question, there is an emerging consensus that conflicting results (Mattingly, 2015; Madsen and Rodgers, 2015) may be a consequence of methodological lapses such as ignoring both industry and other differences (Russo and Fouts, 1997; Crifo and Mottis, 2016; Isaksson and Woodside, 2016; Saeidi *et al.*, 2015).

Such lapses represent a gap in the research literature although there are exceptions that appear to avoid them:

1. Hart and Ahuja (1996) split their complete sample and obtained materially different results (1996: 33 and 36), which imply that the FP-EP relationship varies with environmental performance, although this variation has not yet been subject to thorough investigation.
2. Single industry studies focused on electric utilities (Hughes, 2000), and on the U.S. pulp and paper industry (Clarkson *et al.*, 2004) point to a positive correlation between FP and EP.
3. A study of the pulp paper, chemicals, oil and gas, and metals and mining industries (Clarkson *et al.*, 2011) focused on firms with, respectively, improved or deteriorating environmental performance. They reported that firms with high financial performance later improved their environmental performance, adding: "It would appear that only firms with sufficient financial resources and management capabilities can pursue proactive environmental strategies" (p. 142). Again, these results imply a positive FP-EP relationship.

Our research attempts to explain why the literature on the relation between Financial Performance of firms and their Environmental Performance has provided conflicting results. We do so by improving the specification of the EP – FP model to: include important control variables; to address serial correlation of the regression errors; by reducing the bias in testing the relation at different levels of the dependent variable (Environmental Performance); and by including both long term and short term financial performance as independent variables.

2. THEORY DEVELOPMENT

2.1. Environmental Performance: Dependent Variable

Environmental performance (EP) is an empirical notion that has been operationalized as reduced toxic releases. Toxic releases (lbs) (TR) is the dependent variable in this study. It is a readily available output measure reported to the EPA, facility by facility, and available to researchers in the Environmental Protection Agency (EPA) toxic release inventory (TRI) database. TRI was used by Chatterji, Levine, and Toffel (2009),ⁱ Delmas and Blass (2010), as well as by Clarkson *et al.* (2004) and Clarkson *et al.* (2011). Like these authors, we aggregated the toxic releases of each firm's separate facilities to the firm level. The natural log of TR lbs/\$ Sales is the measure for our dependent variable and controls for scale and technology differences that are substantial across the cases in this study. TR is, of course, an inverse measure of environmental performance – smaller TR means better EP.

2.2. Short and Long-Term Financial Performance: Independent Variables

There is a substantial imbalance in the literature favoring a positive correlation between FP and EP that could be a by-product of bias introduced by researchers or the review process. However, as Sterne and Harbord (2004) explained in their paper on funnel plots and bias, similar imbalances are often a consequence of (unrecognized) heterogeneity across the many samples studied. And, as noted above, the literature has given very little attention to heterogeneity of any kind when the FP-EP relationship is studied.ⁱⁱ

In addition to uncertainties about the FP-EP relationship that might be explained by further attention to heterogeneity, we noted that different measures of financial performance have been used in different studies. Some focused on short-run returns, others on long-term shareholder value (Buysse and Verbeke, 2003; Hart and Ahuja, 1996; Judge and Douglas, 1998). Some called for rigor and recommended using both as we did. We adopted earnings before interest and taxes (EBIT/Sales) as our short-term measure and market value/book value as the long-term measure. The first is an operating measure largely unaffected by the firm's capital structure. The second focuses on the value of its equity, a departure from common practice to facilitate interpretation.ⁱⁱⁱ

2.3. Slack: A Control

Slack is one of several control variables included to ensure we do not overstate the strength of the relationship between our two primary variables. Omitted variables can bias the estimates of the included explanatory variables. Bourgeois (1981: 30) defined organizational slack as a cushion of actual or potential resources that allows an organization to adapt successfully. Later Bourgeois and Singh (1983) identified three types of slack distinguished one from another by ease of access. Consistent with them, we employed 'current ratio (current assets/current liabilities)' for available slack, 'debt/equity' for potential slack, and 'selling, general, and administrative expense (SGA)/sales' for recoverable slack. We mean-centered each measure by industry to accommodate to the realities of the different practices of the three 2-digit industries in this study.

2.4 Other Control Variables

This study also incorporated the following additional controls: capital expenditure, diversification of product line, and firm size. Hart and Ahuja (1996) observed that U.S. manufacturers were spending about 20% of their capital expenditures (named CAPX for convenience) on emissions compliance so we included CAPX divided by net property plant and equipment in this study as a control. It measures the rate at which a firm is renewing its production capabilities.

Some argue that as firms diversify, management pays more attention to the reputation of every operating unit (Dixon-Fowler *et al.*, 2009; Fombrun, 1996) and so they expected diversified firms to strive for better EP. On the other hand, others argue that with diversity, governance issues, control problems, and the cost of coordination rise and that, as a consequence, EP might suffer (Rawley, 2010; Teece, 1980). We measured diversification as the number of 3-digit industries in which every firm participates each year (Wernerfelt and Montgomery, 1988). Firms in this study were active in a minimum of 1 and a maximum of 11 industries.

Firm size seems likely to shape the EP-FP relationship (Hillman and Keim, 2001; Waddock and Graves, 1997). Large firms are likely to have discretionary resources to mitigate environmental problems while small firms do not (Bowen, 2002). Our study measures firm size as the 4th root of Firm Assets to reduce heteroscedasticity.

Table 1 presents the variables included in this study, the measures used, and their means, standard deviations, medians, and ranges. Note that the meta-study by Margolis *et al.* (2007) documents that in corporate social responsibility (CSR) and EP research, the numbers of variables studied was usually small suggesting that misspecification could be shaping results in the literature.

Table 1: Descriptive Statistics

Variable	Measure	N	Mean	Median	sd	Max	Min
EP	Ln TR/Sales	2187	3.73	3.96	2.84	11.25	-13.73
CAPX	CAPX/NetPPE	2267	0.18	0.17	0.10	2.13	0.01
Diversity	No. of 3 digit SIC's	2272	2.43	2.00	1.78	11.00	1.00
Size	4 th √Assets	2272	8.54	7.89	3.37	29.89	3.09
Available Slack	Current Ratio	2151	0.02	-0.16	1.03	9.69	-1.71
Potential Slack	Debt/Equity	2161	-0.01	-0.19	0.72	4.08	-1.26
Recoverable Slack	SGA/Sales	2179	0.00	-0.03	0.12	0.51	-0.21
Short Term FP	EBIT/Sales	2272	0.11	0.10	0.10	0.57	-1.74
Long Term FP	Market Value/ Book Value	2272	2.76	2.32	1.49	8.24	0.75

3. HYPOTHESIS

Hart and Ahuja (1996: 34) concluded that reduced TR was correlated with profitability for firms with high emissions but not for firms with low, implying variation in the FP-EP relationship with respect to EP across and possibly within industries. Hence, our null hypothesis: the FP-EP relationship is invariant with the level of EP.

4. THE FP-EP MODEL

To test the null hypothesis that the FP-EP relationship does not change with the level of EP, we used the following model:

$$TR_{i,t} = F(\beta_i FP_{i,t} + \beta_k \text{Control Variable}_{i,t} + e_{i,t}) \dots \dots (1)$$

where TR = Log-transformed Total Toxic Releases/Sales in t year

FP= Short Term FP (EBIT/Sales) and Long Term FP (Market/Book Value) and

the Control Variables = Slack [Available, Potential, and Recoverable]; and the firm attributes [CAPX, Diversification, and Size ($4^{\text{th}} \sqrt{\text{Assets}}$)].

To ensure a proper test of our hypothesis, we ascertained whether the FP-EP relationship was serially correlated and differed across industries before we determined whether and how the relationship varies with the level of EP.

5. DATA

For this longitudinal study, we integrated data from the EPA TRI for EP and Compustat North America (CAN) for FP, both short and long term, as well as for our Slack measures, and CAPX. Our research context was three 2-digit SIC industries which represented a wide range of polluters. They are all subject to government regulation and social pressure to manage their environmental impacts. They employ a wide array of technologies and have different degrees of dependence on energy. Efforts to improve EP should be normal practice in these industries and, so, bring FP-EP into high relief.

The data series are annual 1991 to 2008 for SIC 31, 32 and 33. SIC 31 includes food/beverage/tobacco, textiles, apparel, and leather. Industry 31 includes General Mills and Sara Lee, companies with limited toxic releases. SIC 32 is paper, printing, publishing, petroleum, chemical, plastics and rubber, and stone/clay/glass/cement. It includes Chevron and Sonoco (a packaging solutions company) and many heavy polluters. These are companies that should be sensitive to pressure for pollution abatement. SIC 33 is primary metals, fabricated metals, machinery, computers and electronic products, electric equipment, transportation equipment, furniture, and miscellaneous manufacturing. Industry 33 includes Alcoa and Ford.

Note that we reduced the sample from approximately 2,100 cases (firm years) to about 1,940 by trimming particular cases (that is single firm years) where the long-term FP was outside its 5% upper and lower limits. Preliminary analysis indicated several seemingly 'implausible values' for this variable. Rather than deleting firms and attempting to reconcile the odd years with 'normal' trends, we applied this simple criterion to allow us to focus on the central core of the data.

6. STATISTICAL METHODS AND RESULTS

Our objective was to test whether the FP-EP relationship varies with EP. To complete this test, we explored the FP-EP relationship within the complete sample, a pool of our three 2-digit industries, as well as separately within each industry. And, as well as completing this test, we hoped to identify plausible methodological explanations for the conflict that characterizes the literature. Our approach was different from most prior research. We included a large number and variety of controls (three types of slack, capital expenditure, diversity of the product line, firm size).^{iv} We deliberately explored serial correlation. We tested for industry differences. And, finally, we used simultaneous quantile regression to directly address the research question (Koenker and Hallock, 2001; Cameron and Trivedi, 2009).

6.1. Setting the Stage

Our first step was to establish a rigorous basis for addressing our research question by exploring serial correlation in the FP-EP relationship in our complete sample and in each of its three industries. To justify this step, we argue that some industries may be able to adjust quickly to, say, changes in technology. Others may take longer. Practice also suggests those with good EP in one year are likely to deliver similar performance in subsequent years. Similarly, firms with poor EP might find it difficult to improve, and, of necessity, follow a partial adjustment model over several more years to lift their performance. Some might simply comply with the law.

Wooldridge tests confirmed that the residuals are serially correlated (Wooldridge, 2013). Here, because the data are an unbalanced panel and because the time-scope of the data is limited, we were unable to determine an optimal lag structure. Nevertheless, tests revealed that the estimated β of the first lag of the dependent variable was consistently significant suggesting a first order process in the error term. Adding the first lag of the dependent variable (DVL1) to the model is an accepted strategy for dealing with autocorrelation. If time was omitted from the estimate, the EP-FP relation would be misspecified, biasing the estimates of the standard errors of the β coefficients, an outcome likely to lead to incorrect inferences about significance.

Thus, after a Hausman Test, we ran fixed effects models using robust standard errors with the first lag of the dependent variable (DVL1) for the complete sample and the three industries. There were no significant results for either of the FP

variables but the addition of DVL1 led to a substantial improvement in the models' explanatory power. The R² statistics for Industry 31 rose from .0009 without DVL1 to 0.43 when DVL1 was included; from 0.0001 to 0 .80 for Industry 32; and, for Industry 33 from 0.0492 to 0.59.

The estimated β coefficients and p-values for DVL1 ranged from 0.315 (0.017) for Industry 31 to 0.691 (.000) for Industry 32. Since each was < 1, the models appear to be dynamically stable although we could not rule out a higher order autoregressive process in the error term. We experimented with DVL2 and longer lags without achieving any improvement in the model's performance.

Tests of differences between the β coefficients of each pair of industries indicated that Industry 31 was different from 32, and 32 different from 33 (Paternoster *et al.*, 1998). We could not show Industry 31 as different from 33. A subsequent Wald test, an exclusion test for the joint significance of all main effect and interaction variables (specifically industry dummies), showed these as 'significant,' implying the three industry models differ one from another.^v

6.2. Testing by Level of Toxic Releases

Quantile regression allowed us to directly test the hypothesis that the FP-EP relationship is invariant with the level of EP. Although not widely used in the strategic management or CSR literature, quantile regression is especially suited to determining whether the FP-EP relationship varies with the level of the dependent variable or does not (Koenker and Hallock, 2001).

By using the full sample, quantile regression allowed us to avoid partitioning the sample on the dependent variable and the bias that would engender. Instead of dividing the sample, every observation was used to estimate the median of each quantile. The optimization process used minimizes the absolute value of deviations from each quantile's median. Overestimates and underestimates are given different weights according to the quantiles chosen, thereby including every observation in each quantile estimate.^{vi} Our results are reported in Table 2. They are specific to the 0.2, 0.4, 0.6, and 0.8 quantiles of the dependent variable, EP, and are a real test of our hypothesis.

Table 2: Estimate of the EP-FP Relationship

Simultaneous Quantile Regression (with Lagged Dependent Variable, DVL1) (Significant results at p ≤ 0.05 in bold)									
EP(Ln TR/Sales)	Complete Sample		Industry 31		Industry 32		Industry 33		
	Coef.	p	Coef.	p	Coef.	p	Coef.	p	
q20									
DVL1	1.002	0.000	0.985	0.000	1.019	0.000	0.980	0.000	
Short Term FP	-0.340	0.399	0.354	0.824	-0.532	0.304	-1.067	0.255	
Long Term FP	0.001	0.979	-0.007	0.930	-0.003	0.881	-0.025	0.443	
Available Slack	0.063	0.064	0.352	0.278	0.032	0.630	0.061	0.400	
Potential Slack	0.050	0.083	0.052	0.539	0.023	0.657	0.093	0.081	
Recoverable Slack	-0.066	0.778	0.405	0.774	-0.016	0.939	0.096	0.863	
CAPX	-1.102	0.000	-1.788	0.235	-0.319	0.469	-1.294	0.136	
Diversity	0.031	0.000	-0.055	0.781	0.029	0.001	0.020	0.356	
Size	0.020	0.001	0.034	0.430	0.014	0.021	0.041	0.021	
_cons	-0.538	0.000	-0.590	0.193	-0.591	0.000	-0.535	0.003	
q40									
DVL1	0.989	0.000	0.990	0.000	0.997	0.000	0.982	0.000	
Short Term FP	-0.059	0.773	0.838	0.522	-0.241	0.392	-0.483	0.306	
Long Term FP	-0.013	0.153	-0.005	0.922	-0.023	0.033	0.008	0.684	
Available Slack	0.031	0.213	0.129	0.501	0.023	0.552	0.036	0.081	
Potential Slack	0.036	0.128	-0.051	0.570	0.055	0.066	0.012	0.701	
Recoverable Slack	0.176	0.285	0.154	0.845	0.157	0.161	0.351	0.412	
CAPX	-1.04	0.000	-1.555	0.103	-0.251	0.273	-1.192	0.006	

Diversity	0.015	0.024		-0.004	0.938		0.022	0.010		0.005	0.707
Size	0.096	0.208		0.019	0.793		0.005	0.307		0.016	0.269
_cons	-0.068	0.211		-0.112	0.734		-0.154	0.003		-0.067	0.553
EP(Ln TR/Sales)	Complete Sample			Industry 31			Industry 32			Industry 33	
	Coef.	p		Coef.	p		Coef.	p		Coef.	P
q60											
DVL1	0.990	0.000		0.941	0.000		0.989	0.000		0.980	0.000
Short Term FP	0.060	0.751		1.042	0.401		-0.449	0.079		0.291	0.615
Long Term FP	-0.021	0.007		-0.024	0.609		-0.016	0.100		-0.025	0.201
Available Slack	0.025	0.370		0.073	0.573		0.028	0.306		0.021	0.236
Potential Slack	0.030	0.196		-0.000	0.998		0.054	0.070		0.014	0.682
Recoverable Slack	0.276	0.046		-0.410	0.508		0.264	0.039		0.444	0.191
CAPX	-0.569	0.012		-1.532	0.078		0.190	0.595		-0.856	0.002
Diversity	0.011	0.132		0.046	0.947		0.009	0.067		0.007	0.580
Size	0.002	0.610		-0.000	0.994		0.004	0.316		-0.004	0.688
_cons	0.122	0.040		0.365	0.403		0.024	0.759		0.204	0.055
q80											
DVL!	0.951	0.000		0.929	0.000		0.954	0.000		0.954	0.000
Short Term FP	0.379	0.128		2.158	0.227		-0.254	0.456		0.276	0.790
Long Term FP	-0.029	0.014		-0.043	0.547		-0.020	0.092		-0.046	0.152
Available Slack	0.049	0.273		-0.074	0.725		0.027	0.500		0.048	0.419
Potential Slack	0.046	0.060		0.442	0.084		0.053	0.058		0.019	0.721
Recoverable Slack	-0.013	0.933		-1.015	0.398		0.241	0.205		0.033	0.941
CAPX	-0.363	0.283		-0.171	0.907		0.066	0.831		-0.798	0.207
Diversity	0.005	0.490		-0.082	0.379		0.011	0.105		-0.009	0.490
Size	-0.011	0.039		-0.036	0.406		-0.011	0.067		-0.006	0.716
_cons	0.567	0.000		1.012	0.087		0.534	0.000		0.644	0.000
Number of obs	1442			97			751			594	
.20 Pseudo R²	.770			.762			.803			.720	
.40 Pseudo R²	.790			.777			.811			.749	
.60 Pseudo R²	.791			.795			.800			.755	
.80 Pseudo R²	.764			.799			.762			.747	

Although there are few significant results in Table 2, the total set reveals differences across the quantiles, industry by industry. There were no significant results for short-term financial performance at $p \leq 0.05$. There is one significant result at $p \leq 0.05$ for long-run financial performance, a β of -0.023 ($p = 0.033$) in Industry 32 (paper, printing, publishing, petroleum, chemical, plastics and rubber, stone/clay/glass/cement) at Q40. There are three significant results at a marginally significant level ($p \leq 0.100$) for Industry 32: One is for short-term financial performance in Q60 (-0.449 , $p = 0.079$); the others are for long-term financial performance in Q60 (-0.016 , $p = 0.100$) and Q80 (-0.020 , $p = 0.092$). As Table 2 makes clear, the number of significant results is sparse and varies across industries although we note that each of the β estimates referred to above has the expected negative sign implying a positive FP-EP relationship, albeit of limited reliability.

Here we want to bring two results to the reader's attention because they have methodological implications for published research in this field. Note that in addition to the industry specific results reported above, there were two apparently

significant results for long-run FP in the complete sample: an estimated β of -0.021 ($p = .007$) at Q60 and another of -0.029 ($p = .014$) at Q80. These two “apparently significant results” are of *some* moment because they echo results reported where heterogeneity was not tested. They suggest that failures to test for heterogeneity may be one cause of some of the conflict in the literature. Had we limited the research to the complete sample and not tested for industry differences, we might have simply reported that market value was negatively related to toxic releases although “only for the worse performing companies,” and so was positively related to EP. In that circumstance, our results would have lent credence to those segments of the literature reporting a positive FP-EP correlation for heavy polluters.

Recall that we added DVL1 to the model to allow proper estimation by addressing serial correlation in the residuals. The size of the β s for the DVL1s in Table 2 are, however, ≈ 1.0 , suggesting caution. Normally high β for the lagged dependent variable suggest a risk of overestimating the significance of the theoretical variables. Here, because we have few significant results to report, there is little risk of overestimating significance, but caution suggests that the FP-EP relationship could be even weaker than Table 2 suggests.

7. DISCUSSION

Why is the FP-EP relationship so difficult to measure? Ignoring the complete sample results, as is appropriate when industry differences have been documented, Table 2 shows that although we estimated a total of 24 short- and long-term financial performance coefficients (for 2 variables, 3 industries, and 4 quantiles), only one was significant. The one significant result for long-term financial performance in Industry 32 at Q40 was an exception, a telling result because of the steps taken to ensure rigor throughout this research. This one significant result in 24 tests (1/24) suggests that the FP-EP relationship as specified and controlled (here) is of limited strength. At best, outside of our research context, it seems, financial performance is likely to be ‘found’ significantly related to environmental performance only in very narrowly defined contexts. Subsequently, we estimated equation 1 for eight different 4-digit SIC industries and can only report results similar to those above. The results of Table 2 appear to be robust.

Thus, we conclude there is no compelling case for a strong FP-EP relationship; the evidence here is insufficient for rejection of the null hypothesis.^{vii} The results suggest that to establish the actual form of the FP-EP relationship, if it exists, we need a richer and more tightly focused database possibly including different measures of pollution and toxicity. If the relationship ultimately proves to be very complex, considerably more advanced and complex methods will be required to define it. It is possible, of course, that an FP-EP relationship may be found at the plant level (Dutt and King, 2014) although obtaining relevant financial data for a cross industry is likely to be very difficult, since such data are proprietary.

Our research also attests to the fact that the effects of time and differences across and within industries cannot be addressed piecemeal. While the DVL1 coefficients are quite similar across industries 31, 32, and 33, that's not our main concern. In Table 2, recoverable slack is significant at Q60 but only for Industry 32. At Q60, CAPX has a significant coefficient but only for Industry 33. At Q20 and Q40, diversity has significant coefficients for Industry 32 but not Industries 31 and 33. Size has significant coefficients in Industries 32 and 33 at Q20 but not in Industry 31. At Q80 no control variable is significant. All these results establish that for Q20, Q40, and Q60, there are industry differences although at Q80 there is no such evidence. Only at Q40 is there evidence that the effect of long-term FP is different for Industry 32 compared to either 31 or 33. Clearly, however, the specified FP-EP relationship varies across industries and quantiles.

In light of this study, therefore, researchers addressing the FP-EP relationship should be aware that if they do not first address relevant lag effects and deal with performance differences across industries and quantiles, it is unlikely they will be able to explain industry environmental performance differences and throw more light on the nature of the FP-EP relationship.

8. LIMITATIONS

We were unable to reject the null hypothesis “there is no FP-EP relationship.” Yet our results consistently reveal coefficient differences across the quantiles for both financial performance and the controls. Thus, we have provided evidence suggesting that the FP-EP relationship is likely to vary with environmental performance, as Hart and Ahuja (1996) reported and as Hughes (2000), Clarkson *et al.* (2004), and, later, Isaksson and Woodside (2016) posit.

Thoroughly dealing with autocorrelation and industry differences complicates the task of establishing the nature of the FP-EP relationship, yet it is necessary. Of course, research to establish the power of different variables to explain EP demands data, and larger databases inevitably introduce heterogeneity to the study.

Hughes (2000) and Clarkson *et al.* (2004) eliminated much of the potential confusion related to industry differences by studying single industries and partitioning them insightfully. Hughes took advantage of a “natural experiment” by separating utilities that were explicitly identified in the Clean Air Act of 1990 from the rest. Clarkson *et al.* (2004) simply looked at the top 50% versus the bottom 50% on a TR/COGS index. Clarkson *et al.* (2011) separated firms from four

polluting industries into ‘improving, regressing, and stable’ environmental performance groups. In that paper, further degrees of heterogeneity were held to be of unimportant.

9. CONCLUSION

Do the effects of FP on EP shift with the level of EP? The direct answer is they probably do although establishing the true effect is likely to be difficult to establish because it seems likely to differ in specific contexts. This study suggests, however, that the conflict in the literature as to whether financial and environmental performance are positively, negatively, or unrelated should be tabled: the true nature of the FP-EP relationship is elusive. Moreover, note that in our complete sample studies results matched the many of the conflicting FP-EP results published over the past thirty years –. But as we hope is clear, these multi-industry results were invalidated by our three separate industry studies. With our complete sample and three specific industry results in mind, the conjectures of Russo and Fouts (1997) about methodological lapses being a likely cause of the equivocal results reported in the literature, then, as now, seem likely to be well founded.

REFERENCES

- Bourgeois, L. J. (1981). On the measurement of organizational slack. *Academy of Management Review*, 29-39.
- Bourgeois, L. J., Singh, J. V. (1983). Organizational slack and political behavior within top management groups. Faculty of Management Studies, University of Toronto.
- Bowen, F. E. (2002). Does size matter? *Business & Society* 41 (1), 118.
- Buyse, K., Verbeke, A. (2003). Proactive environmental strategies: a stakeholder management perspective. *Strategic Management Journal* 24 (5), 453-470.
- Cameron, A. C., Trivedi, P. K. (2009). *Microeconometrics using stata*: Stata Press.
- Chatterji, A. K., Levine, D. I., Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility?. *Journal of Economics & Management Strategy* 18 (1), 125-169.
- Clarkson, P., Li, Y., Richardson, G. (2004). The market valuation of environmental capital expenditures by pulp and paper companies. *The Accounting Review* 79 (2), 329-353.
- Clarkson, P., Li, Y., Richardson, G., Vasvari, F. (2011). Does it really pay to be green? Determinants and consequences of proactive environmental strategies. *Journal of Accounting and Public Policy* 30, 122-144.
- Crifo, P., Mottis, N. (2016). Socially responsible investment in France. *Business & Society* 55 (4), 576-593.
- Delmas, M. A., Blass, V. D. (2010). Measuring corporate environmental performance: the trade-offs of sustainability ratings. *Business Strategy and the Environment* 19 (4), 245-260.
- Dixon-Fowler, H. R., Slater, D. J., Johnson, J. L., Ellstrand, A. E. (2009). Beyond "Does it pay to be green?" A meta-analysis of moderators of the CEP and CFP relationship. *Academy of Management Annual Meeting*.
- Dutt, N., King, A. A. (2014). The judgment of garbage: end-of-pipe treatment and waste reduction. *Management Science* 60 (7), 1812-1828.
- Fombrun, C. J. (1996). *Reputation*. Harvard Business School Press, Boston, MA.
- Hart, S. L., Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between pollution prevention and firm performance. *Business Strategy and the Environment* 5, 30-37.
- Hillman, A. J., Keim, G. D. (2001). Shareholder value, stakeholder management, and social issues: what's the bottom line? *Strategic Management Journal* 22 (2), 125-139.
- Hughes, K. E. (2000). The value relevance of nonfinancial measures of air pollution in the electric utility industry. *The Accounting Review* 75 (2), 209-228.
- Isaksson, L. E., Woodside, A. G. (2016). Modeling firm heterogeneity in corporate social performance and financial performance. *Journal of Business Research* 69, 3285–3314.
- Judge, W. Q., Douglas, T. J. (1998). Performance implications of incorporating natural environmental issues into the strategic planning process: an empirical assessment. *Journal of Management Studies* 35 (2), 241-262.
- Koenker, R., Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives* 15 (4), 143–156.
- Madsen, P., Rodgers, Z. J. (2015). Looking good by doing good: the antecedents and consequences of stakeholder attention to corporate disaster relief. *Strategic Management Journal* 36, 776-794.

Margolis, J. D., Elfenbein, H. A., Walsh, J. P. (2009). Does it pay to be good...and does it matter? A meta-analysis of the relationship between corporate social and financial performance. <https://ssrn.com/abstract=1866371> or <http://dx.doi.org/10.2139/ssrn.1866371> (accessed 01/11/17).

Mattingly, J. E. (2015). Corporate social performance: a review of empirical research examining the corporation-society relationship using Kinder, Lydenberg, Domini social ratings data. *Business & Society* 54, 1-44.

Paternoster, R., Brame, R., Mazerolle, P., Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology* 36 (4), 859-866.

Rawley, E. (2010). Diversification, coordination costs, and organizational rigidity: evidence from microdata. *Strategic Management Journal* 31 (8), 873-891.

Rost, K., Ehrmann, T. (2015). Reporting biases in positive research paradigms in management: the example of win-win corporate social responsibility. *Business & Society*. First Published Feb 25, 1-49 (0007650315572858).

Russo, M. V., Fouts, P. A. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal* 40 (3), 534-559.

Saeidi, S. P, Sofian, S., Saeidi, P., Saeidi, S. P., Saeidi, S. A. (2015). How does corporate social responsibility contribute to firm financial performance? The mediating role of competitive advantage, reputation, and customer satisfaction. *Journal of Business Research* 68, 341-350.

Salinger, M. A. (1984). Tobin's Q, unionization, and the concentration-profits relationship. *The RAND Journal of Economics* 15 (2), 159-170.

Sterne, J. A. C., Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal* 4 (2), 127-141.

Teece, D. J. (1980). Economies of scope and the scope of the enterprise. *Journal of Economic Behavior and Organization* 1 (3), 223-247.

Toffel, M. W., Marshall, J. D. (2004). Improving environmental assessment: a comparative analysis of weighing methods used to evaluate chemical release inventories. *Journal of Industrial Ecology* 8 (4), 143-172.

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

Wernerfelt, B., Montgomery, C. A. (1988). Tobin's q and the importance of focus in firm performance. *The American Economic Review* 78 (1), 246-250.

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach, fifth ed.* Mason, OH: South-Western Cengage Learning.

End Notes:

¹ Toffel and Marshall (2004) criticized TR as measure of pollution explaining that the TRI data set is not verified by any authority despite its being maintained by the EPA. They added that TR lbs may be too simple a measure for some purposes, elaborating that pollution can be also characterized by its human or ecosystem toxicity, by its cumulative effects, and by the medium of release. Such detail seems best suited to the study of pollution within specific geographic areas or at specific plant locations such as the Dutt and King (2014) study of the adoption of 'End of Pipe' waste treatment at manufacturing plants and later on-going waste reduction. Dutt and King (2014: 1815) also provide additional insight into the limitations of the TRI data

² Rost, K, and Ehrmann, T. (2015), Reporting Biases in Positive Research Paradigms in Management: The Example of Win-Win Corporate Social Responsibility, *Business & Society*. First Published Feb 25, pp. 1-49 (0007650315572858)

³ Many studies have used one of the several versions of Tobin's Q as a long term measure. But most formulations of Tobin's Q include a replacement value making the results difficult to interpret. For Salinger (1984) Tobin's q = Market Value (Equity + Debt + Preferred Stock) divided by what he called Replacement Value (the sum of property, plant and equipment less accumulated reserves for depreciation, plus Short Term Assets). Note that Replacement Value as defined above has little to do with replacement. Take the specific example of the Georgia Gulf (GG) Corporation. GG was spun off by Georgia Pacific on December 31 1984 with Fixed Assets having a 'Face' value of \$165MM against a prior 'Book' value on Georgia Pacific's books of approximately \$365MM. Yet, at the time of a second public offering in 1986, the replacement value of these same assets was stated as being about \$1billion (Goldman Sachs, 1986). Given that several alternate formulations of Tobin's Q are used by researchers, users must pay close attention to the particular measure used.

⁴ Our initial research plan was to include macroeconomic conditions and policy with variables such as a % real GDP growth and energy prices. However, preliminary research showed that these variables are essentially Fixed Effects, and we cannot use them in models with Fixed or Random Effects because they are co-linear with the binary variables for firms and years. If one uses the Stata command *xtsum* (listing the shared environmental variables), it should show no zero values for the standard deviations of these variables, except for the entity and time variables such as id or year. In this research, the standard deviations of these macro variables are not zero, and not even particularly low, but they should be. For example, GDP Growth, a macro variable with the same value for all firms, is co-linear with a binary variable that expresses the intercept.

⁵ Note that Clarkson et al (2011) is a cross industry study anchored to an *a priori* partitioning of the data. In that study there are 5 categories of firms ranked by EP across time. This research strategy imposes a substantial structure on the data and, in this case, without tests of the degree to which other sources of heterogeneity affects the study. Clarkson et al (2011) Tables 4 and 5 indicate no adjustment for heterogeneity of variance of the error term. It appears that after partitioning the sample by 5 categories, they estimated the equations with all 5 categories in one pool.

⁶ Koenker and Hallock (2001: 145) write “...Just as we can define the sample mean as the solution to the problem of minimizing a sum of squared residuals, we can define the median as the solution to the problem of minimizing a sum of absolute residuals. The symmetry of the piecewise linear absolute value function implies that the minimization of the sum of absolute residuals must equate the number of positive and negative residuals, thus assuring that there are the same number of observations above and below the median. What about the other quantiles? Since the symmetry of the absolute value yields the median, perhaps minimizing a sum of asymmetrically weighted absolute residuals—simply giving differing weights to positive and negative residuals—would yield the quantiles. This is indeed the case. Solving $\min \sum p_r (y_i - \epsilon) \in$ where the function $p_r(\cdot)$ is the tilted absolute value function ...that yields the r^{th} sample quantile as its solution.

⁷ At that stage of the research, neither FP measure was significant at $p < 0.05$ at or even less rigorous standards ($0.05 < p < 0.10$) in any of the four regressions