DOES SOCIAL PERFORMANCE REALLY LEAD TO FINANCIAL PERFORMANCE? ACCOUNTING FOR ENDOGENEITY.

Roberto Garcia-Castro
IESE Business School
Avda. Pearson 21
08034 Barcelona-Spain
Tel.: (34) 932534200
docrgarcia@iese.edu

Miguel A. Ariño
IESE Business School
Avda. Pearson 21
08034 Barcelona-Spain
Tel.: (34) 932534200
MAArino@iese.edu

Miguel A. Canela
University of Barcelona
Gran Via de les Corts Catalanes, 585
08007 Barcelona - Spain
Tel.: (34) 934021636
miguelcanela@ub.edu

November, 2007

ABSTRACT

The empirical relationship between a firm's social performance and its financial performance is still not well established in the literature. Despite more than 30 years of research and more than 100 empirical studies on the issue, the results are still mixed. We argue that the heterogeneous results found in previous studies are not due exclusively to problems related with the measurement instruments or the samples used. Instead, we posit that a more fundamental problem related with the endogeneity of social strategic decisions could be driving most of the empirical findings. We show, using a panel data of 658 firms from 1991-2005, how some of the results found in previous research change and some are even reversed when endogeneity is taken properly into account.

Keywords: Social Performance; Financial Performance; Stakeholder Management; Corporate Social Responsibility; Endogeneity

What is the relationship between the social performance of firms –the quality of the relationships between a firm and its stakeholders— and their financial performance? Over the last 35 years numerous researchers have tried to provide a definitive and clear answer to this fundamental question for both academics and managers. The results of these previous studies on the relationship between *social performance* (hereinafter, SP) and either market- or accounting-based measures of *financial performance* (hereinafter, FP) have been mixed (Ullmann, 1985; Griffin and Mahon, 1997; Roman, Hayibor and Agle, 1999; Margolis and Walsh, 2001, 2003; Post, Preston and Sachs, 2002). Some scholars have found a positive relationship (e.g., Cochran and Wood, 1984; McGuire, Sundgren and Schneeweis, 1988; Coffrey and Fryxell, 1991; Waddock and Graves, 1997b; Berman et al., 1999; Hillman and Keim, 2001). Other works find more ambiguous or negative relationships (e.g., Alexander and Buchholz, 1978; Aupperle, Carroll and Hatfield, 1985; McWilliams and Siegel, 2000).

In the most comprehensive survey performed to date on the link between SP and FP, Margolis and Walsh (2001; 2003) review 127 studies published in articles and books since the early work of Moskowitz (1972). In 109 of the 127 studies, SP has been treated as the independent variable, predicting FP. Margolis and Walsh (2003) conclude that out of these 109 studies, one half (54) pointed towards a positive SP-FP relationship, 20 showed mixed results and 28 studies reported non-significant relationships. Only 7 studies showed a negative relationship (Margolis and Walsh, 2003: 278).

Despite the fact that a majority of previous research tends to support a positive SP-FP link and although recent, more refined meta-analyses of past research on the issue have shown a dominance of a positive relationship between SP and FP (Orlitzky,

Schmidt and Rynes, 2003), such a relationship is still far from being well established in the literature. Post et al. (2002) summarize the empirical evidence found so far in the field:

> "The safest generalization from them [empirical studies on the link between social and financial performance] is that the empirical evidence on this matter is somewhat unreliable and the results mixed. However it is important to note that *there is very little evidence of a negative association between social and financial performance*…To put it another way, the empirical studies do not prove that corporations can "do well by doing good", but neither do they disprove that view, and there is no substantial evidence that corporations can "do well by doing harm" (Post, Preston and Sachs, 2002: 28). *Emphasis added.*

How can we explain the heterogeneity of these findings? Is it possible to generalize the positive link between SP and FP found in the majority of previous works? Does such a positive link hold in the long and also in the short run? In the literature, we find all kinds of explanations related with sampling problems, issues related with the validity and reliability of SP and FP measures, omission of relevant controls, mediating mechanisms, or the lack of a causal theory (Margolis and Walsh, 2003). While most of those problems are endemic in most of the strategic management research, there are three in particular that are especially relevant in SP-FP research.

First, the heterogeneity of the findings could be due to a lack of consistent and reliable instruments to measure SP (Waddock and Graves, 1997a; 1997b). Previous studies have used corporate reputation indexes, or distributed questionnaires to measure the firm's commitment to certain stakeholders. More recent studies have tried to refine SP measures and use more consistent and comparable across-study measures such as the Kinder, Lydenberg, Domini (KLD) index of SP (see table 1). However, although the

more recent studies that use KLD to measure SP tend to support a positive relationship between SP and FP (Hillman and Keim 2001, Waddock and Graves, 1997b, Berman et al., 1999), some contradictory results are still found (McWilliams and Siegel, 2000).

Insert table 1

A second explanation for the heterogeneity of empirical results is that SP and FP may have a relationship that changes with circumstances which may not yet be understood well enough to be embodied in control variables (Preston and Post, 1975; Waddock and Graves, 1997a; McWillliams and Siegel, 2000).

Finally, a third aspect that must be better understood in SP-FP research is the relationship between short- and long-run performance. Cross-sectional empirical studies tend to measure both SP and FP the same single year, and therefore, the long-term consequences of certain decisions affecting stakeholders are left unexplored. Introducing the short- and long-term dimension can contribute to explaining the inconsistency of previous empirical findings. For example, one interesting study testing the relation between the short and long term is Ogden and Watson (1999). In a longitudinal study of 10 water supply companies operating in the UK, these two authors found that whereas a high SP[1] had a negative impact on firm's *current profitability* –as managers typically had to incur in certain expenses in order to attend to the needs of certain stakeholders —, it also had a significant long-run positive effect on *shareholders' returns.* Similar conclusions to Ogden and Watson (1999) are reached by

[1] In their study, Odgen and Watson (1999) adopt a narrow view of SP and include only customer welfare.

García-Castro, Ariño and Canela (2006), using a wider sample of firms and a time horizon of 7 years.

Despite the importance of using a consistent measure of SP, including all relevant control variables and distinguishing between short- and long-run financial effects, we assert in this paper that there is a more powerful reason for the heterogeneity of previous findings that may affect all of them. We assert in this paper that the decision of top management to improve a firm's SP (i.e., decisions oriented to improving the quality of the relationships between the firm and its stakeholders) is *endogenous*. Such a decision is likely to be correlated with *unobserved firm-specific variables* such as the organization's culture, the quality of its top management, decision-making style, management's ethical attitudes and values or any other hard to observe variables. As a matter of fact, recent research finds evidence for such a correlation between the CEO's values –often difficult to observe or measure— and socially-oriented firm policies and strategies (Agle, Mitchell and Sonnenfeld, 1999). Insofar as the CEO's values can also affect firm performance, endogeneity problems will be biasing all the estimates obtained in the equations.

This paper deals with the specific issue of the *endogeneity* of strategic decisions. We do so in the context of the literature that studies the relationship between SP and FP. We are not aware of any prior work in this particular literature that addresses this issue explicitly –with the sole exception of McWilliams and Siegel (2000)[2]. This paper proceeds as follows. In the next section, we explain the endogeneity problem in this particular field of research. Next, we present a longitudinal dataset that allows us to apply recent econometric methods to deal with the problem of endogeneity following

[2] However, McWilliams and Siegel (2000) limit their discussion to issues related with the proper, complete specification of the econometric model without fully discussing the problems and challenges posited by the existence of endogeneity.

previous works in other related fields (Shaver, 1998; Campa and Kedia, 2002; Hamilton and Nickerson, 2003; Villalonga, 2004). A discussion of the results found and some conclusions close the paper.

ENDOGENOUS STRATEGIC DECISIONS

The basic endogeneity problem

The endogeneity problem is well-known, and is typically taken into account in fields such as economics where econometric techniques exist to correct --at least partially-- for endogeneity (Heckman, 1974; Greene, 1993). However the use of these econometric techniques in strategic management research has so far been limited (Hamilton and Nickerson, 2003).

The basic problem lies in that managers make strategic decisions not randomly – a standard assumption in many cross-sectional regression models— but based on expectations on how their choices will affect future performance (Hamilton and Nickerson, 2003). These expectations arise from internal factors that managers, presumably, know very well but are difficult to observe by external researchers (e.g., firm culture, internal configuration of capabilities, CEO's personal values,…). The problem arises because any statistical analysis that does not take into account these unobserved variables (if they are not included in the model's specification as control variables) can suffer from biased coefficient estimates. The biases result from omitted variables correlated with both the strategic decision and firm performance (Hamilton and Nickerson, 2003; Wooldrigde, 2002)

Previous research has shown that both the direction and the size of the bias can have important consequences, leading, in extreme cases, to radically opposite

conclusions. For instance, in a recent paper, Campa and Kedia (2002) show that the conclusion as to whether there is a diversification discount or not can be reversed when the endogeneity of the diversification decision is taken into account. Campa and Kedia (2002) show that the diversification *discount,* extensively documented in previous empirical studies, becomes a *premium* when proper methodological controls for endogeneity are introduced. In a similar paper, Villalonga (2004) reaches a similar conclusion. Other works in different fields have shown the critical impact of endogeneity, proposing alternative ways to deal with it (Masten, 1996; Shaver, 1998).

Endogeneity problems in the social issues in management (SIM) research

In the particular field of SIM, and more specifically in SP-FP research, the heterogeneity in the conclusions shown by previous findings could be suggesting that endogeneity is a relevant issue. Besides, a large majority of previous studies tend to find a *positive* effect of SP on FP (Waddock and Graves, 1997b; Margolis and Walsh, 2001; Post et al., 2002). However, it is a matter of common observation and common sense that if firms are to satisfy all the multiple stakeholders (employees, customers, community, suppliers,…), they may sometimes have to sacrifice financial results…at least in the short run. Despite the obviousness of this fact, how is it that previous studies systematically fail to find empirical evidence of a negative effect of SP on FP?[3] As a matter of fact, these previous studies typically measure FP in the short run (e.g., Waddock and Graves, 1997b; Berman et al., 1999).

In SP-FP research, few works have tried to deal with the endogeneity problem. McWilliams and Siegel (2000) have been among the first authors to point out the

[3] As explained in the introduction of this paper, the empirical evidence found so far for a *negative* effect of SP on FP is merely anecdotic, as has been acknowledged by several scholars (Ullmann, 1985; Orlitzky, Schmidt and Rynes, 2003; Margolis and Walsh, 2001, 2003; Post, Preston and Sachs, 2002).

methodological flaws in previous SIM research. They argued "that the positive and significant coefficient on CSP [SP], as reported by Waddock and Graves (1997), could simply reflect the impact of R&D intensity on firm performance [if correlation (R&D intensity, CSP)>0[4]]. It is impossible to isolate the impact of CSP on firm performance unless the model is properly specified. A similar argument could be made for other omitted regressors…if they are also positively correlated with CSP and firm performance." McWilliams and Siegel (2000: 606).

In fact, McWilliams and Siegel show how the sign of the regression coefficient of SP (on firm performance) found in previous studies changed—from positive coefficients to neutral—when R&D intensity is specified in the equation. This is a paradigmatic example of how endogeneity affects the results obtained. One of the dimensions of SP in Waddock and Graves (1997b) study is the relationship with customers. But, in fact, the relationships with customers are likely to be correlated with R&D and the introduction of new products, innovation etc. And McWilliams and Siegel show R&D intensity to be positively correlated with both SP and FP (McWilliams and Siegel, 2000: 608). As a consequence, the positive effect between SP and FP found by Waddock and Graves (1997b) is overestimated.

Other authors in the SIM field have found other variables to be correlated with SIM decisions. For example, Agle, Mitchell and Sonnenfeld (1999) found some evidence that CEO's values can influence the adoption of socially-oriented policies affecting employees or customers. However, if these *omitted* variables are typically, by their very nature, hard to observe and measure, how can we make sure that our econometric model is not underspecified? Is it possible to take into account all the

[4] McWilliams and Siegel (2000) make a compelling argument for the link between CSP, R&D intensity and firm performance. In addition, they argue that something similar to R&D intensity happens with advertising intensity.

relevant variables that influence a manager's decision in a given firm? The answer is, of course, no. However, in recent years some solutions to this problem have been posited. One alternative is the use of longitudinal/panel data. A second approach consists of using instrumental variables in the model's specification.

Longitudinal/panel data

Having the possibility of gathering panel data on strategic variables and FP, including observable control variables, allows the researcher to better estimate how a firm performs under different strategic regimes. What follows is based in the work of Hamilton and Nickerson (2003). Further explanations and more details can be found there.

Let's consider a case where a firm chooses between having a positive socially-responsible policy with its stakeholders ($\mathbf{SP_{it}}$=1 if the firm i chooses strategy $\mathbf{SP_1}$ in time period t) or not having a pro-social policy at all ($\mathbf{SP_{it}}$=0 if the firm i chooses strategy $\mathbf{SP_0}$ in time period t). *Under some assumptions*[5] about the nature of the endogeneity problem, we can specify the econometric model as:

$$\pi_{it} = \gamma SP_{it} + X_{it}\beta + \theta_i + \zeta_{it} \qquad \mathbf{(1)}$$

Where π_{it} is the financial performance outcome; $\mathbf{X_{it}}$ includes the control variables observable by the researcher; θ_i is a time-invariant error term; and ζ_{it} is a time-

[5] In particular, in equation **(1)** we are assuming: (I) The unobservable variables that affect performance under $\mathbf{SP_1}$ are assumed to be the same as those influencing performance under $\mathbf{SP_0}$. (II) The error term is assumed to consist of a time-invariant, firm-specific component, θ_i and a time-varying component, ζ_{it}, that is uncorrelated across periods, so that $\varepsilon = \theta_i + \zeta_{it}$, where ε is the total error. (III) The only omitted variables that affect strategy choice (SP) and performance **(π)** *do not change over time* (Hamilton and Nickerson, 2003).

varying error term. As usual, **i** is the subscript for each individual and **t** is the subscript for time.

The main difficulty for researchers is the fact that the firm-specific error θ_i is not directly observable, but it does significantly affect the regression coefficients of the other explanatory variables. In terms of strategic management, θ_i represents all those attributes of a firm that distinguishes it from the rest and make it somehow unique. Quite clearly, it is not the same that firm or manager A implements policy $\mathbf{SP_1}$ or that firm or manager B implements the same policy. As we argued above, taking into account and correcting for the error induced by θ_i is especially relevant in the presence of elusive variables such as trust, values, culture, etc., all of them variables hard to measure by its very nature.

One problem with standard OLS estimation –or the GLS estimation used for random effects models—is that it assumes that the error term is uncorrelated with the observed covariates, $\mathbf{SP_{it}}$ and $\mathbf{X_{it}}$. This specification rules out the existence of firm-specific unobserved factors that affect both the strategic decision **(SP)** and financial performance (π), which is precisely the endogeneity problem. Hence, estimates of **SP** (γ) will be biased upwards or downwards.

One solution to the problem of standard OLS estimation is to specify a fixed-effects model, allowing θ_i to be correlated with $\mathbf{SP_{it}}$ and $\mathbf{X_{it}}$ (as we think is the case in real firms). Fixed effects are incorporated by either including a set of firm indicator variables into the regression (we have different intercepts for each individual in the sample), or differentiating equation **(1)** in order to eliminate θ_i:

$$\pi_{it}-\pi_{it-1}= \gamma(SP_{it}-SP_{it-1}) + (X_{it}-X_{it-1})\beta + (\zeta_{it}-\zeta_{it-1}) \qquad (2)$$

Estimating **(2)** via OLS yields a consistent estimate of the effect of γ since θ_i is now out of the regression, if assumptions I, II and III hold, of course. Unfortunately, these assumptions do not always hold. Hamilton and Nickerson (2003: 71) point out that equation **(2)** may not be appropriate when: (1) the value of γ is different for different groups of firms; (2) when the effect of θ_i on performance is different under $\mathbf{SP_1}$ and $\mathbf{SP_0}$; (3) when the changes in strategy **(SP)** are not exogenous (why does a firm or its managers change its strategy during the panel period?). If some of the problems outlined are likely to be present in the sample, then additional tools are needed to account for endogeneity. One proposed solution is the use of instrumental variables, a solution considerably more complex than panel data fixed estimation.

Instrumental variables

As we mentioned above our basic hypothesis is that firms that choose to engage in social activities are not a random sample of firms. In principle, it is possible to determine **SP** in terms of a set of variables that influence **SP** but are *not correlated* with FP. Specifically we can assume that the **SP** of a firm i in time t is given by

$$\mathbf{SP_{it}} = \beta \mathbf{Z_{it}} + \mu_{it} \qquad \textbf{(3)}$$

where $\mathbf{Z_{it}}$ is a set of firm characteristics that affect $\mathbf{SP_{it}}$ but *are uncorrelated* with π_{it} in equation **(1)** and μ_{it} is an error term. $\mathbf{Z_{it}}$ are called instrumental variables or instruments because they allow us to estimate the effect of SP on performance indirectly, without using directly our original endogenous measure of SP.

The main difficulty with the instrumental variables estimation is how to identify valid instruments because most of the observable firm characteristics are already included in the main performance equation, causing the system to be unidentified (Campa and Kedia, 2002). For this reason, the characteristics of a good instrument are such that it is *not correlated* with the error term in the main performance equation and also, that it *is correlated* with the endogenous variable of interest, in our case, $\mathbf{SP_{it.}}$ If these instruments are available, the estimation of equation **(1)** using the instrumental variable equation **(3)** would yield unbiased estimators of $\boldsymbol{\gamma}$ (more details on this in Hamilton and Nickerson, 2003; Campa and Kedia, 2002; Shaver, 1998).

RESEARCH METHODS

Sample and data collection

Previous studies have used standard OLS regression analysis in order to test the hypothesis that SP has a positive impact on FP (Waddock and Graves, 1997b; McWilliams and Siegel, 2000; Hillman and Keim, 2001). In this study, we have used a panel based on the 658 firms included in KLD database[6]. KLD is an independent rating agency specialized in the assessment of corporate social performance across a range of dimensions related to stakeholder concerns. In total, the panel covers 15 years (1991-2005). Financial data as well as firm-level control variables were collected from *Datastream*.

Although KLD covers more than 3,000 firms since 2003, only 658 firms have been covered during the entire period since 1991. Consequently, we have restricted our panel to those 658 firms for which historical data are available. The 658 firms are all

[6] Kinder, Lydenberg, Domini & Co. Inc., 129 Mt. Auburn St, Cambridge, MA 02138, USA.

US-based. KLD covers most of the firms listed in S&P500 and Domini 400 social index. Given that most of the companies listed in these indexes are covered by KLD, we do not expect to have any sort of sample selection bias in the sample used for this study as that is the population of firms we want to study here. Also, common method bias does not affect our study as the data for the independent and dependent variables were collected from two completely different sources.

Estimation methods

The resulting panel was unbalanced. The firm is the primary stratification variable, so that there is a 658-item unbalanced panel with a time series between one and fifteen observations in each stratum. The panel includes variables that are time-varying for the panel period, such as KLD measures, FP and most of the control variables, and time invariant variables, such as industry. We combine OLS, fixed-effects models and instrumental variable estimation with the purpose of comparing our results with previous findings and also to account for endogeneity.

Measures

Dependent variables

In order to homologate our results with previous findings, we use the following 4 measures of FP: ROE, ROA, Tobin's Q and MVA. Those 4 measures were among the measures of performance used most often in the past (Margolis and Walsh, 2001). Consistently with previous studies, ROE is calculated as net income over total equity. ROA is calculated as operating income over total assets. For Tobin's Q, we use the

proxy of market-to-book value ratio. Previous studies have used the market-to-book value as approximations to Tobin's Q (Huselid, 1995; Wiggins and Rueffli, 2002).

We operationalize shareholder value creation using the market value-added or MVA (Stern Stewart, 1996). MVA was calculated as follows:

MVA= Market value – Capital;

MVA= market value-added
Market value = Firm's market value or market capitalization of the firm
Capital = book value of *equity* and *debt* invested in the firm

MVA is the difference between the market value of a company (both equity and debt) and the capital that lenders and shareholders have entrusted to it over the years in the form of loans, retained earnings and paid-in capital. If MVA is negative, the company has destroyed wealth (Stern Stewart, 1996). Although the MVA has received some criticisms as a measure of shareholder value creation (Fernández, 2002) it is still a widely used proxy for value creation, and it has been used in recent studies in strategic management (Hillman and Keim, 2001). MVA provides some advantages over other traditional measures of firm performance as, for instance, accounting measures of performance (ROE, ROA), that are typically more short-term oriented measures of performance (Hayes and Abernathy, 1980) and may be subject to manipulation by managers (McGuire et al., 1988).

Independent variables

Social Performance. We use the Kinder, Lydenberg, Domini (KLD) index as a measure of the quality of stakeholder relations. KLD measures have been used in previous research to study the link between SP and FP in premier management journals (Hillman and Keim, 2001; Waddock and Graves, 1997b; Agle, Mitchell and Sonnenfeld, 1999; Berman et al., 1999; McWilliams and Siegel, 2000; Graves and Waddock, 2000; Coombs and Gilley, 2005). The KLD rating makes several advances beyond those used in earlier research as it constitutes a multi-dimensional measure of SP, consistently measured by a group of professionals with the same criteria across a large sample of firms and where different information sources are triangulated in order to find out the final score for each firm (Waddock and Graves, 1997b). A main advantage of using the KLD rating is that it is publicly available information, and thus it allows researchers, in different studies, to compare their findings using the same measurement instruments.

Hence, following previous studies, the items chosen in our study came from 5 categories of the KLD instrument: employee relations, customer/product issues, community relations, diversity issues and environmental issues (Waddock and Graves, 1997b; Hillman and Keim, 2001). Each of the 5 KLD categories is the aggregate of the different attributes considered by KLD. Appendix 1 provides details on the factors used in determining ratings for each of the five categories. KLD measurement of SM combines quantitative criteria with expert judgment consistently applied across the pool of firms in order to determine whether a firm has strengths or concerns for each of the factors depicted in Appendix 1. According to KLD methodology, a number "1" in any category means a *strength* or *concern* of that firm for that particular category whereas a

"0" indicates that a company did not meet the required criteria for the strength or the concern.

We constructed a unique index that captured the quality of the stakeholder relations for each firm. Given the difficulties in arriving at a universal ranking of priorities among stakeholders (Mitchell et al., 1997), we follow Hillman and Keim's (2001) procedure of giving equal weight to each of the 5 categories of the SM construct. After adding up all the 5 items, the resulting KLD index ranged --in our sample of 658 firms-- from a minimum score of -8 to a maximum of +12, with the average score being 0.747 and the standard deviation being 2.517. These statistics were computed for the 658 firms in our sample for all of the years during which the firm is present in the panel, giving a total of 7,541 firm-year observations for the KLD variable. The resulting KLD index is equivalent to previously used KLD indexes in the literature (Waddock and Graves, 1997b; Hillman and Keim, 2001). Interestingly, the fact that firms of high reputation accredited in many case studies and qualitative studies such as Ben & Jerry, Southwest Airlines, Deere & Company, or Medtronic, were placed at the top of the resulting ranking increased our confidence in the validity of the KLD's instrument. Conversely, firms involved in corporate scandals such as WordCom or Tyco appear well at the end of the list in the previous year's rankings.

Control variables

In accordance with previous studies of stakeholder management and firm performance, we control for *size*, *industry* and *risk* effects (Aupperle et al., 1985; Pava and Krausz, 1996; Waddock and Graves, 1997b; Hillman and Keim, 2001; Coombs and Gilley, 2005).

Size is a factor that has been suggested to affect both firm performance and stakeholder orientations. After checking that three measures of size (number of employees, total sales and total assets) were highly correlated, we chose total sales as our control for firm size. *Industry* has been operationalized in this paper by using the FTSE Industrial classification codes obtained from Datastream. We create one dummy variable for each individual industry resulting in a total of 37 dummies. Very few firms --only 13-- changed from one industry to a different one during the panel period. For that reason, we decided to treat industry as a time invariant variable[7]. *Firm risk* has been operationalized using company's *beta* as reported in *Datastream Thomson Financials*. The beta is a measure of market risk which shows the relationship between stock volatility and market volatility. As all of the firms analyzed were US firms, we did not have to use country-specific control variables.

In addition, McWilliams and Siegel (2000) show that other firm-specific variables such as *R&D intensity* may affect both KLD and performance and therefore suggest that it also be included as a control variable. We include it in the models operationalizing R&D intensity as R&D expenses over sales, consistently with McWilliams and Siegel (2000). Finally, we control for the *leverage ratio*, also included in previous SP-FP research (Waddock and Graves, 1997b). We operationalize the leverage ratio as total debt over equity.

Model specification

The baseline model is an OLS pooled cross-sectional estimation with the following specification:

[7] In those rare cases where a firm is in two different industries for different years we only consider the industry to which the firm has belonged for the highest number of years over the 15-year period of our sample.

$$\Pi_{it} = \alpha + \beta_1 KLD_{it} + \beta_2 Risk_{it} + \beta_3 Sales_{it} + \beta_4 R\&D_{it} + \beta_5 leverage_{it} + \beta_6 - \beta_{42}(Industry_j) + \theta_i + \varepsilon_{it} \quad \textbf{(4)}$$

Π_{it} = ROE, ROA, Tobin's Q or MVA of firm i in time t

KLD_{it} = social performance of firm i in time t = Σ(Community relations + Employee relations + Diversity policies + Environmental concern + Product (customer concern) of firm i in time t

$Risk_{it}$ = Beta of firm i in time t

$Sales_{it}$ = Total sales of firm i in time t

$R\&D_{it}$ = R&D expenses over sales of firm i in time t

Leverage=Total debt over total equity of firm i in time t

$Industry_j$ = 37 time-invariant dummy variables

i = 1...658 firms

t = 1991-2005; 15 years

j = 1…37 industries

θ_i is the time-invariant error term and ε_{it} is a time-varying error term.

The estimation of equation (4) is called pooled ordinary least squares because it corresponds to running OLS on the observations pooled across i and t (Wooldridge, 2002). For this estimation we are taking each cross section for each year as if they were independent random samples from the relevant population. Then, we proceed to compare our baseline model (4) with the fixed-effects model (5):

$$\Pi_{it} = \alpha_i + \beta_1 KLD_{it} + \beta_2 Risk_{it} + \beta_3 Sales_{it} + \beta_4 R\&D_{it} + \beta_5 leverage_{it} + \beta_6 - \beta_{42}(Industry_j) + \varepsilon_{it} \quad \textbf{(5)}$$

In equation (5), we introduce a different intercept α_i for each firm (i), controlling in this way for unobserved firm characteristics. Note that in equation (4) we are not accounting for the effect of θ_i on performance, and thus, we expect the coefficients of the other covariates on the right side of the equation to be biased upwards or downwards.

Testing for endogeneity

To test for the existence of endogeneity in our data, we use Hausman's test (see Hausman, 1978). Hausman's test is based on the difference between the random-effects estimator (which is efficient under the null hypothesis of *no endogeneity* and inconsistent under the alternative) and the fixed-effects estimator (which is consistent under both but inefficient under the null).

The application of Hausman's test to our sample produced a non-positive definitive covariance matrix of the differences between the random and the fixed effects, making it impossible to compute the test. In practice, application of Hausman's test involves subtracting the covariance matrices of the random effect and the fixed effect estimator. In general, the resulting covariance matrix will be positive definitive. However, these results only hold asymptotically. For a given fixed sample like ours, the resulting covariance matrix could be non-positive definite. In such a case, it is not possible to compute Hausman's test (Wooldridge, 2002).

Mundlak (1978) suggests an alternative test equivalent to the Hausman's test consisting of estimating, in the same econometric model, the original endogenous and control variables and the mean for each firm of the variables we suspect, a priori, to be endogenous. If the regression coefficients of the mean variables are significant, then endogeneity problems exist in the sample. We applied Mundlak's test using the four dependent variables of interest, ROE, ROA, MVA and Tobin's Q. The four regression coefficients proved to be significant ($p< 0.01$) for the means of the KLD variable. Hence, we can reject the null hypothesis of no endogeneity. This result confirms the relevance of endogeneity in this kind of research and the need to account for endogeneity in our sample.

We account for endogeneity in several ways. We first reproduce the results existing in the literature using OLS and we identify a similar pattern of association between SP and FP in our sample. Under the assumption that all non-observable firm characteristics that lead to endogeneity do not change over time and they have the same impact for different levels of KLD and that changes in KLD levels are exogenous over the course of the panel period, we take advantage of the panel data to perform fixed-effects estimation in order to eliminate θ_i from the analysis and obtain unbiased estimates of γ. Finally, we relax the previous assumptions and perform an instrumental variables estimation to control for endogeneity in these circumstances.

RESULTS

OLS estimation

Descriptive statistics and Pearson correlations between the variables are presented in Table 2. In Table 3, we show the results obtained (GAC) and a comparison

with earlier studies, using KLD as a measure of SP (WG, WS and HK). Previous research used mainly the OLS estimation. Thus, we first reproduce previous methods using pooled cross-sectional OLS in order to identify in our sample the positive link between KLD and FP found in previous studies. We control for the same variables as in previous research and the R-squared obtained are comparable with those obtained by previous researchers (Table 3). We found positive and significant ($p<0.01$) coefficients for KLD in all four equations using the four dependent variables: 1.509, 0.392, 1995 and 0.186 for ROE, ROA, MVA and Tobin's Q respectively. Thus, our OLS results (GAC) confirm previous results (WG, WS and HK), even when we consider over a 15-year time horizon, which indicates that the positive SP-FP link is fairly consistent over time.

Insert table 2

Insert table 3

However, the estimates obtained by OLS may be biased if omitted variables are correlated with performance and KLD. We next use fixed-effects to test whether that is the case in our sample.

Fixed-effects estimation

In the GAC columns in Table 3, we introduce fixed effects to control for unobservable firm characteristics which affect KLD as discussed in the previous

sections (FE columns under GAC). As expected, the F-tests that all θ_i =0 were all significant at the 0.01 level when the four performance variables were used as dependent variables in the fixed-effects estimation, leading us to reject the null hypothesis that θ_i is zero. Hence, this test confirms the existence of non-zero θ_i in our equation, providing thus additional evidence that pooled cross-sectional regression produces biased estimates of the effect of KLD on performance.

As Table 3 shows, the introduction of fixed effects in the models reduces the KLD coefficient to 0.618 and not significant in the case of the ROA, to 0.125 and only significant at the 0.10 level for the ROA, to -384 and not significant for the MVA and to -0.132 and significant at the 0.10 level for Tobin's Q.

The introduction of firm fixed effects reduces interfirm variability in the data used for estimation and it might increase the noise-to-signal ratio in the estimation (Campa and Kedia, 2002). However the signs of the coefficients in all other control variables, with the only exception of R&D intensity in the case of Tobin's Q, remain identical to those from the OLS estimation. The only coefficient that changes its sign from the OLS to the FE estimation is KLD. Altogether, the results confirm that there are unobserved variables driving the positive link between KLD and FP found in previous studies (and in the OLS columns under GAC in table 3).

We also try different specifications of the fixed-effects model including firm and year effects in the same equation and the results are consistent with those shown in Table 3 (GAC, FE), where only firm effects are included for comparative purposes with previous literature. Unfortunately, fixed-effects estimation has its own drawbacks. In particular, in the FE estimation we are assuming that all the unobserved variables that affect KLD and performance simultaneously do not change over time and that the

changes in KLD are exogenous (Hamilton and Nickerson, 2003), which are problematic assumptions in this case. When these assumptions are relaxed, we need to use instrumental variables in order to estimate an unbiased coefficient for KLD.

Instrumental variables (IV) estimation

As we discussed in the methods sections, the most difficult issue in the IV estimation is the ability to produce valid instruments in order to estimate equation **(3)**. Two important characteristics of a good instrument are that 1) it reasonably predicts the endogenous variable (KLD in our model) and 2) it is not correlated with the dependent variable in our main model (ROE, ROA, Tobin's Q, MVA). Previous research has shown that a firm's SP is significantly influenced by the industry the firm belongs to, the firm's visibility, its size, the specific characteristics of stakeholder groups in terms of power, legitimacy and urgency, and also by corporate governance (CG) (Mitchell, Agle and Wood, 1997; Rehbein, Waddock and Graves, 2004; Brammer and Millington, 2006; Eesley and Lenox, 2006). All of these variables could qualify as instruments for SP but, given that size is often correlated with FP, it cannot be used as a valid IV. In addition, we do not have specific data regarding stakeholder characteristics such as power, legitimacy and urgency for the firms in our sample and thus we cannot use them in the IV estimation.

We identified three sets of valid instruments. The first set consists of industry characteristics. Previous studies document the influence of industry on the SP (Rehbein et al., 2004; Brammer and Millington, 2006). Some industries, such as food, textiles and apparel, receive more external pressure for product-related concerns while others, such as refining, rubber, plastic, telephony or utilities, for instance, receive more external

pressure for their environmental and energy practices (Rehbein et al., 2004: 261). As industry is correlated with our performance measures, we redefine our dependent variable: we define FP for our four variables (ROE, ROA, Tobin's Q, MVA) as the value relative to the average of the industry for a given year. Such a measure is by construction independent of any observable characteristic that affects the FP of all firms in a given industry and year in the same manner (Campa and Kedia, 2002). After redefining the performance variables, industry dummies qualify as valid instruments as they significantly predict SP and they are uncorrelated with firm performance by construction.

The second set of instruments is related to CG. Earlier works have emphasized that CG conditions may have an effect on the SP of firms (Brammer and Millington, 2006). We found three concrete attributes of CG, uncorrelated with performance, that predict SP. The first is limited executive compensation (LEC). Firms with a moderate compensation for its top management and board members tend to follow the so-called codes of good governance and tend to engage in socially responsible behaviours. LEC is operationalized as firms with a total compensation of less than $500,000 per year for a CEO or less than $30,000 per year for outside directors. The second variable is ownership strength (OWS). A company that owns between 20% and 50% of another company that KLD has cited as having an area of social strength or is more than 20% owned by a firm that KLD has rated as having social strengths, is more likely to show a high SP. OWS may drive the decision to engage in pro-social policies to the extent that socially responsible owners give effect to their participation in the firm's capital through the appointment of socially responsible managers and also influencing the decisions made by the management team. The last variable is transparency (TRS),

which measures a firm's effectiveness in reporting social and environmental performance measures. All other things being equal, firms with more external pressure for social and environmental reporting will be more likely to engage in pro-social policies.

The third instrument is related with firm visibility, which is supposed to externally influence SP. In the absence of a finer-grained measure of visibility, we operationalize it in this study with a dummy variable indicating whether a firm is listed in the Standard & Poor's 500 or not (SP500). Firms listed in SP500 are supposed to have a higher exposure to investors, media, activists, etc., and therefore they are expected to have higher visibility.

Altogether, the three sets of instruments are good predictors of KLD while at the same time are determined mainly exogenously since a firm happens to be in an industry at a certain point in time (i.e., it is not a strategic decision made each year). The same thing happens with SP500 and the CG variables --LEC, OWS and TRS-- that are also induced to a large extent by external legal requirements and the ownership characteristics of corporations.

Next, we use these three sets of instruments to estimate SP in equation **(3):**

$$\mathbf{KLD_{it}} = \alpha + \beta_1 \mathbf{LEC_{it}} + \beta_2 \mathbf{OWS_{it}} + \beta_3 \mathbf{TRS_{it}} + \beta_4 \mathbf{SP500_{it}} + \beta_5 - \beta_{41} (\mathbf{industry_j}) + \mu_{it} \quad \mathbf{(6)}$$

In Table 4, we present the results of the OLS estimation that allows us to estimate the levels of KLD in a firm. We will use them later in these OLS estimates of KLD to perform the IV regression in a second step. The resulting model accounts for 24 percent of the variance found in the KLD variable. As expected, most of the explanatory variables in equation **(6)** were shown to significantly affect the levels of KLD. Thirty of

the industry dummies, TRS and SP500 had a significant impact on KLD. Only LEC (p= 0.841) and OWS (p= 0.647) were not significant at conventional levels.

Insert table 4

Next, we perform an instrumental variable regression in a second step. In Table 5, we estimate again the impact of KLD on ROE, ROA, MVA and Tobin's Q using an IV approach. We instrument KLD by using the variables in the model shown in Table 4 above: LEC, OWS, TRS, industry dummies and SP500. The results we show in Table 5 are consistent with our initial hypothesis and the results found using fixed-effects estimation: the coefficients for ROE, ROA, MVA and Tobin's Q change from positive to negative (for ROE and MVA) or they stop being positive and significant (for ROA and Tobin's Q), suggesting again that endogeneity issues are affecting the relationship between KLD and FP at the firm level.

Insert table 5

DISCUSSION

The results found in the previous section indicate that KLD does not impact performance per se. The positive impact found in previous studies (Waddock and Graves, 1997b; Berman et al., 1999; Graves and Waddock, 2000; Hillman and Keim, 2001) is due mainly to the fact that the firms that adopted high standards of KLD self-selected themselves. That positive effect dilutes when endogeneity is properly taken into account. We have found that companies with certain characteristics (e.g., good

management quality, certain values, a certain culture, etc.) are the ones more likely to adopt KLD practices and these unobserved firm characteristics are driving performance. These findings suggest that future research should look at the firm-specific characteristics (through clinical research) that push firms to adopt those KLD practices in the first place. Only when we understand the reasons behind KLD adoption by managers, will we be able to establish the logical cause-and-effect connection between SP and FP.

Our findings also suggest a critical examination of the KLD measurement of SP. We conjecture that the KLD methodology and other analogous social rating systems for firms may be missing a critical dimension related with the quality of management (Ghoshal and Barlett, 1994; Waddock and Graves, 1997a) underlying the adoption of these policies. Although the correctness of the previous statement should be confirmed or disconfirmed by further research (probably of the clinical research type), future methodology developments to measure a firm's SP should look more carefully at the management quality dimension if they are to improve their predictive power in terms of FP in the long and the short run.

Now we turn our attention to some of the limitations of the present paper. While the construction of the 15-year panel data is the most complete panel for the study of SP-FP to date, it contains large amounts of historical data on firms that are often hard to gather, consequently giving rise to unbalanced panel data with missing values. Whether these missing values are affecting the results found in this paper is difficult to ascertain but it is still a caveat to be taken into account. However, the fixed-effects estimation method is quite robust and the results found are expected to hold if the same methods are applied to other samples or other countries. The IV estimation could be improved

substantially in the future if we are able to come up with better predictive models of KLD levels. A good example of how this could be accomplished is the recent paper by Eesley and Lenox (2006) where they explore secondary stakeholder groups and the conditions under which those groups are more likely to elicit positive social responses by the corporations being targeted. Eesley and Lenox (2006) build a unique database containing stakeholder actions. Databases of this nature, if generalized to all firms in the SP500, for example, could be excellent instruments for predicting KLD in subsequent empirical works.

Finally, although the evidence presented tends to support our initial endogeneity hypothesis, the results were different for alternative measures of FP and thus future research should carefully select the most appropriate dependent variable when studying social issues in management (time horizon to be used, accounting or market-based measures,…). The same applies to the independent variable: we used KLD as a proxy for SP. Whether the results we found hold also when alternative proxies are used (e.g., SAM social and environmental score) is an empirical question worth addressing in future research.

CONCLUSION

We have argued along this paper that endogeneity problems have plagued previous research studying the SP-FP link. While endogeneity problems are not unique to the SIM field we have argued and empirically shown that ignoring them can significantly bias the estimates of the relation between SP and FP. More specifically, we have shown in this paper, using the most complete panel data available using KLD (1991-2005), that the positive relationship found in most of the previous research on the

link between SP and FP becomes a non-significant or even a negative relationship when endogeneity is properly taken into account.

The implications of this finding are twofold. First, there is an obvious methodological consequence: given the magnitude of the bias induced by the endogeneity and self-selection problems shown in this paper, it is urgent that researchers come up with new and improved ways of dealing with them. In this paper, we have used fixed-effects and instrumental variable estimations as a first approach to the problem, following previous works in other related fields (Shaver, 1998; Campa and Kedia, 2002; Hamilton and Nickerson, 2003; Villalonga, 2004). Future papers should be able to refine some of the instrumental variables suggested in the paper while also providing alternative and innovative ways to deal with the endogeneity of social strategic decisions.

There is a second implication related with the measurement of SP. The possible negative relationship found between SP (KLD) and FP in the fixed-effects and IV estimation *do not* suggest that such relationship is negative: it is simply suggesting that there are unobserved variables correlated with both SP (KLD) and FP that mediate the SP-FP relationship. This clarification is critical and it urges us to reflect on the nature of SP indexes and measurement tools, what are the underlying dimensions they are not able to capture and what are the ways in which we can improve those indexes and measurements in such a way that they really capture the essence of what SP in a business context is. We believe that a systematic aspect missing in SP indexes like KLD or SAM (Sustainable Asset Management) is a sound measure of the quality of management (Ghoshal and Barlett, 1994; Waddock and Graves, 1997a)[8]. The quality of

[8] There is not a unique definition of management quality When we use the term management quality in this paper, we have in mind a concept along the lines of Ghoshal and Barlett (1994), Waddock and

management is likely to be related with high levels of KLD but also with FP, as good, responsible managers will also deliver financial results on top of satisfying the different corporate constituencies apart from the shareholders. This observation was made by Waddock and Graves (1997b: 315):

> "It is thus entirely possible that there are direct linkages between the overall quality of management and CSP [SP]…Further, if quality of management is a critical variable in financial outcomes, as the relationships identified in this study suggest, then controlling for the quality of management while assessing the CSP-financial performance link might also be beneficial".

Despite the fact that this observation was made more than ten years ago, very little progress has been made since then to advance our knowledge about the interaction between quality of management and SP. Our paper suggests that such an interaction may be not only important for theoretical purposes but that it may also affect empirical findings.

REFERENCES

Agle BR, Mitchell RK, Sonnenfeld JA. 1999. Who matters to CEOS? An investigation of stakeholder attributes and salience, corporate performance and CEO values. *Academy of Management Journa*l **42** (5): 507-525.

Alexander GJ, Buchholz RA. 1978. Corporate social responsibility and stock market performance. *Academy of Management Journal* **21** (3): 479-486.

Aupperle KE, Carroll AB, Hatfield JD. 1985. An empirical examination of the relationship between corporate social responsibility and profitability. *Academy of Management Journal* **28** (2): 446-463.

Graves (1997a) or even Perez-Lopez (1993), that emphasizes motivational and organizational aspects of good managers over strictly technical (product-market) aspects.

Berman SL, Wicks AC, Kotha S, Jones TM. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and the firm financial performance. *Academy of Management Journal* **42** (5): 488-506.
Brammer S, Millington A. 2006. Firm size, organizational visibility and corporate philanthropy: an empirical analysis. *Business Ethics: A European Review* **15** (1): 6-18.
Campa JM, Kedia S. 2002. Explaining the diversification discount. *Journal of Finance* **57** (4): 1731-1762.
Cochran PL, Wood RA. 1984. Corporate social responsibility and financial performance. *Academy of Management Journal* **27**: 42-56.
Coffey BS, Fryxell GE. 1991. Institutional Ownership of Stock and Dimensions of Corporate Social Performance: An Empirical Examination. *Journal of Business Ethics* **10** (6): 437-444.
Collins JC, Porras JI. 1994. *Built to last: successful habits of visionary companies*. New York: HarperBusiness.
Coombs JE, Gilley KM. 2005. Stakeholder management as a predictor of CEO compensation: Main effects and interactions with financial performance. *Strategic Management Journal* **26**: 827-840.
Eesley C, Lenox MJ. 2006. Firm responses to secondary stakeholder action. S*trategic Management Journal* **27**: 765-781.
Fernández P. 2002. *Valuation methods and shareholder value creation*. Academic Press: San Diego CA.
Garcia-Castro R, Canela MA, Ariño MA. 2006. Over the long run? The impact of stakeholder management on short run and long run shareholder value creation. *Mimeo.*
Ghoshal S, Bartlett CA. 1994. Linking organizational context and managerial action: the dimensions of quality of management. *Strategic Management Journal* **15** (5): 91-112.
Graves SB, Waddock SA. 2000. Beyond built to last…stakeholder relations in built to last companies. *Business and Society Review* **105**: 393-418.
Greene WH. 1993. *Econometric analysis*. 2nd edition. New York: Macmillan.
Griffin JJ, Mahon JF. 1997. The corporate social performance and corporate financial performance debate. *Business & Society* **36** (1).
Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1** (1): 51-78.
Hausman J. 1978. Specification tests in econometrics. *Econometrica* **46**: 1251-1271.
Hayes R, Abernathy WJ. 1980. Managing our way to economic decline. *Harvard Business Review* **58** (4): 67-78.
Heckman J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* **42**: 679-94
Hillman AJ, Keim GD. 2001. Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal* **22** (2): 125-139.
Huselid MA. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal* **38** (3): 635-672.
Margolis J, Walsh J. 2001. *People and profits?* Lawrence Erlbaum Associates: New Jersey.
Margolis J, Walsh J. 2003. Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly* **48**: 268-305.
Masten SE. 1996. Empirical research and transaction cost economic: challenges, progress, directions. In *Transaction cost economics and beyond*, J. Grownewegen (ed). Kluwer: Boston MA; 43-64.
McGuire JB, Sundgren A, Schneeweis T. 1988. Corporate social responsibility and firm financial performance. *Academy of Management Journal* **31**: 854-872.
McWilliams A, Siegel D. 2000. Corporate social responsibility and financial performance: correlation or misspecification? *Strategic Management Journal* **21**: 603-609
Mitchell RK, Agle BR, Wood DJ. 1997. Toward a theory of stakeholder identification and salience: defining the principle of who and what really counts. *Academy of Management Review* **22** (4): 853-886.

Moskowitz M. 1972. Choosing socially responsible stocks. *Business and Society Review* **1**: 71-75.
Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46:** 69-85.
Ogden S,Watson R. 1999. Corporate performance and stakeholder management: balancing shareholder and customer interests in the U.K. privatized water industry. *Academy of Management Journal* **42** (5): 526-538.
Orlitzky M, Schmidt F, Rynes S. 2003. Corporate social and financial performance: a meta-analysis. *Organization Studies* **24**: 403-441.
Pava M, Krausz J. 1996. The association between corporate social responsibility and financial performance: the paradox of social cost. *Journal of Business Ethics* **15**: 321-357.
Perez-Lopez JA. 1993. *Fundamentos de la dirección de empresas*. Ediciones Rialp: Madrid.
Post JE, Preston LE, Sachs S. 2002. Managing the extended enterprise: The new stakeholder view. *California Management Review* **45** (1): 6-28.
Preston LE, Post JE. 1975. Measuring Corporate Responsibility. *Journal of General Management* **2** (3): 45-52.
Rehbein K, Waddock S, Graves SB. 2004. Understanding shareholder activism: which corporations are targeted? *Business & Society* **43** (3): 239-267.
Roman R, Hayibor S, Agle BR 1999. The relationship between social and financial performance. *Business & Society* **38** (1).
Shaver JM. 1998. Accounting for endogeneity when assessing strategy performance; does entry mode choice affect FDI survival? *Management Science* **44** (4): 571-85.
Stern Stewart. 1996. *The Stern Stewart Performance 1000*. Stern Stewart Management Services: New York.
Ullmann AA. 1985. Data in Search of a Theory: A Critical Examination of the Relationships Among Social Performance, Social Disclosure, and Economic Performance of U.S. Firms. *Academy of Management Review* **10** (3): 540-557.
Villalonga B. 2004. Does Diversification Cause the "Diversification Discount"? *Financial Management* **33** (2): 5-27.
Waddock SA, Graves SB. 1997a. Quality of management and quality of stakeholder relations: Are they synonymous? *Business & Society* **36** (3): 250-279.
Waddock SA, Graves SB. 1997b. The corporate social performance – financial performance link. *Strategic Management Journal* **18** (4): 303-319.
Wiggins RR, Rueffli TW. 2002. Sustained competitive advantage: temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science* **13**: 82-105.
Wooldridge JM. 2002. *Econometric analysis of cross section and panel data.* MIT Press: Cambridge MA

Table 1. Previous Empirical Findings. Social performance (KLD) – Financial performance

Study	Waddock and Graves (1997b) WG	Berman, Wicks, Kotha and Jones (1999) BWKJ	Graves and Waddock (2000) GW	Williams and Siegel (2000) WS	Hillman and Keim (2001) HK	García-Castro, Ariño and Canela (2006) GAC
Year	1989-1991	1991-1996	1991-1997	1991-1996 (average)	1994, 1995, 1996	1991-2005
Data	469 US Standard & Poor's 500 firms belonging to 13 4-digit SIC code industries	81 US Fortune 500 firms belonging to different 4-digit SIC code industries	11 pairs of firms from Built to Last (Collins & Porras, 1994)	524 firms	308 US Fortune 1000/Standard & Poor's 500 firms belonging to different 2-digit SIC code industries	658 fims in KLD and Datastream
Financial performance	ROA, ROE, ROS	ROA	ROE, ROA, ROS	Accounting measure	MVA, ROA, ROE, Tobin's Q	ROA, ROE, ROS, MVA, Tobin's Q
Findings	SP (communities and environment) leads to better FP	Positive relationships with employees and customers positively affect FP	Positive relationship over time between SP and FP	The impact on KLD on performance changes with alternative specifications of the model	Stakeholder management is positively associated with shareholder value creation (MVA)	Positive SP-FP relation when standard OLS is used and non-significant or negative when FE or IV estimation are used
Relationship KLD-performance	Positive	Positive	Positive	Neutral	Positive (MVA)	Biased by unobserved firm-specific variables
Method	OLS	Pooled times series model. Two-step GLS	Trend analysis; T-tests.	-OLS?	OLS	OLS, fixed effects and random effects estimations
Long-run	SR: positive LR: -	SR: positive LR: -	SR: positive LR: positive	SR: neutral LR: -	SR: positive LR: -	SR: - LR: -
Endogeneity	No	No	No	No	No	Yes
Sample	Cross-sectional	Longitudinal/panel	Longitudinal	Cross-sectional	Cross-sectional	Longitudinal/Panel
Instrumental variables	No	No	No	No	No	Yes
Observations	469	486	22	524	308	>3000**

* All studies depicted in Table 1 use KLD ratings in order to measure firms' SP
** Firm-year observations

Table 2. Descriptive statistics and Pearson correlation coefficients*

Variable	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11
1 ROE	16.57	52.56											
2 ROA	10.54	8.53	0.30										
3 MVA	3236	41825	0.03	0.22									
4 Tobin's Q	3.86	14.12	0.28	0.19	0.07								
5 KLD	0.75	2.52	0.09	0.10	0.03	0.03							
6 Beta	0.82	1.44	-0.02	0.00	-0.01	0.00	0.03						
7 Size	8075906	16100000	0.01	-0.05	0.01	0.01	0.01	0.02					
8 R&D intensity	0.04	0.07	-0.07	-0.01	0.19	0.06	0.10	0.06	-0.04				
9 Leverage	114.36	777.22	0.07	-0.05	-0.22	0.35	0.02	0.01	0.07	0.00			
10 LEC	-0.31	0.54	-0.05	-0.01	-0.07	-0.07	-0.06	0.01	-0.25	-0.18	-0.03		
11 OWS	-0.01	0.14	0.01	0.02	-0.02	0.00	0.02	-0.02	-0.05	-0.03	0.01	0.03	
12 TRS	0.05	0.23	-0.05	-0.01	0.02	-0.01	0.14	0.03	0.20	0.08	0.04	-0.05	0.00

***** Correlations equal to or greater than 0.03 are significant at $p < 0.05$

Table 3. Comparison of the Effects of KLD on Financial Performance

STUDY	Waddock and Graves (1997b)			Williams and Siegel (2000)		Hillman and Keim (2001)	García-Castro, Ariño and Canela (2006)							
	WG			WS		HK	GAC							
	ROA	ROE	ROS	Accounting measure	Accounting measure	(MVA)	ROE		ROA		MVA[2]		TobinQ	
	OLS	OLS	OLS	OLS	OLS	OLS	OLS[3]	FE	OLS[3]	FE	OLS[3]	FE	OLS[3]	FE
KLD	.024***	.081	.021**	.141***	-.062	.128**	1.509***	.618	.392***	.125*	1995***	-384	.186***	-.132*
Beta	No	No	No	Yes	Yes	.041	-.363	-.779	.085	.015	-339.876	-1536.04	-.016	-.104
Size	-.502E-6*	.136E-6	-.427E-6	Yes	Yes	-.202**	2.25e-8	4.83e-9	-2.47e-8***	-3.27e-8***	.0003***	.001	-4.23e-9	-9.87e-9
Industry Dummies	Yes	Yes	Yes	No	Yes	Yes	Yes	n/a[1]	Yes	n/a[1]	Yes	n/a[1]	Yes	n/a[1]
R&D intensity	No	No	No	No	.263***	No	-55.038**	-221.407***	-20.610***	-125.758***	47492***	68565.50	5.125	-13.805**
Leverage	-.120***	-.471***	-.115***	No	No	No	.007***	.006***	.0001	.0002*	-1.399**	-.199	.006***	.006***
R^2	.29	.07	.20	—	—	.42	.09	.02	.23	.13	.20	.06	.21	.15
Adjusted R^2	.27	.04	.17	.10	.29	.41	.08		.22		.19		.20	
F-statistic	11.55***	2.20***	6.99***	—	—	35.132***	8.97***	12.85***	27.73***	93.66***	19.17***	35.16***	20.65***	90.20***
No. of observations (firm-year obs.)	**469**	**469**	**469**	**524**	**524**	**308**	**(3334)**	**(3334)**	**(3462)**	**(3462)**	**(2928)**	**(2928)**	**(2920)**	**(2920)**

*p< 0.10; **p< 0.05; ***p< 0.01

[1] Very few firms –only 13—changed from one industry to a different one during the panel period and for that reason we decided to treat industry as a time invariant variable. Consequently, industry dummies do not apply to the case of fixed-effects models as only time-varying variables can be estimated in those models.

[2] MVA is measured in $ Million.

[3] Although we use the notation "OLS", in the four OLS models in GAC we are performing a pooled cross-sectional OLS estimation.

Table 4. OLS estimates for KLD

	Coefficient	Standard error
LEC	0.018	0.088
OWS	0.184	0.402
TRS	2.275***	0.192
SP500 (3 lag)	-0.213*	0.124
Industry dummies[1]		
R^2	0.24	
F-statistic	22.73***	
Observations	2974	

*p< 0.10; **p< 0.05; ***p< 0.01
LEC: Limited executive compensation
OWS: Ownership strength
TRS: Transparency in social and environmental reporting
SP500: dummy variable. "1" if the company is listed in the S&P500 index, otherwise, "0".
[1]A total of 37 dummies representing 37 different industries were introduced in the model.

Table 5. Instrumental variable (IV) estimation

	ROE[1]	**ROA**[1]	**MVA**[1]	**Tobin's Q**[1]
	IV	**IV**	**IV**	**IV**
KLD[2]	-0.059	0.148	-27.149	0.067
Beta	-0.274	-0.066	601.996	0.450
Size	4.77e-8	-9.50e-9	0.0002	-6.79e-9
R&D intensity	-16.340	-3.876	35367.060	3.247
Leverage	0.006	0.00004	-1.467	0.005
R^2	0.02	0.02	0.03	0.21
Adjusted R^2	0.02	0.02	0.03	0.20
F-statistic	6.15***	0.83***	10.63***	85.29***
No. of observations (firm-year obs.)	**1656**	**1750**	**1677**	**1677**

*p< 0.10; **p< 0.05; ***p< 0.01

[1]The difference between the firm's performance for each year and the average of the industry it belongs to for each year is used as the dependent variable for ROE, ROA, MVA and Tobin's Q, respectively.

[2] KLD has been instrumented using the variables in the model shown in Table 4 above: LEC, OWS, TRS, industry dummies and SP500.

Appendix 1
KLD ratings data. Inclusive social rating criteria

Strengths	Concerns
Product	
• Quality	• Product Safety
• R&D/Innovation	• Marketing/Contracting Controversies
• Benefits the Economically Disadvantaged	• Antitrust Disputes
• Other Strength	• Other Concern
Environment	
• Clean Energy	• Hazardous Waste
• Beneficial Products & Services	• Regulatory Problems
• Pollution Prevention	• Ozone Depleting Chemicals
• Recycling	• Substantial Emissions
• Other Strength	• Agricultural Chemicals
	• Climate Change
	• Other Concern
Employee Relations	
• Cash Profit Sharing	• Union Relations Concern
• Employee Involvement	• Health and Safety Concern
• Health and Safety Strength	• Workforce Reductions
• Retirement Benefits Strength	• Retirement Benefits Concern
• Union Relations Strength	• Other Concern
• Other Strength	
Community	
• Charitable Giving Strength	• Negative Economic Impact
• Innovative Giving	• Investment
• Non-US Charitable Giving	• Controversies
• Support for Housing	• Tax Disputes
• Support for Education	• Other Concern
• Volunteer Programs	
• Other Strength	
Diversity	
• CEO	• Controversies
• Promotion	• Non-Representation
• Board of Directors	• Ownership Concern
• Work/Life Benefits	• Other Concern
• Women & Minority Contracting	
• Employment of the Disabled	
• Gay & Lesbian Policies	
• Other Strength	