The relevance of Corporate Social Responsibility criteria to explaining firm profitability: A case study of the publishers of the Dow Jones Sustainability Indexes[1] [2]

Cristiana Mănescu[3] and Cătălin Stărică[4]

First draft: May 2007

This version: June 2007

---

[1] *This research was conducted with funding from the Swedish research programme Behavioral Impediments to Sustainable Investment, funded by the Foundation for Strategic Environmental Research, MISTRA.*

[2] The views expressed in this paper are those of the authors and do not necessarily represent those of Sustainable Asset Management (SAM) Group. All the corporate indicators in the present study have been used at the sole discretion of the authors of this article based on original data provided by SAM Group. Any possible error in the interpretation or manipulation of such data remains the sole responsibility of the authors.

[3] Center for Finance, School of Economics, Business Administration and Law, Gothenburg, Sweden. Email: cristiana.manescu@cff.gu.se

[4] Department of Economics and Department of Mathematical Statistics, Gothenburg University and Chalmers University of Technology, Gothenburg, Sweden. Email: starica@math.chalmers.se Web page: www.math.chalmers.se/~starica

Abstract

This study conducts an in-depth analysis of the impact of Corporate Social Responsibility (CSR) practices as captured by an unique set of scores from the Sustainable Asset Management (SAM) Group[5] on firm profitability measured by the Return on Assets (ROA). The fourteen SAM scores cover all three main aspects of the Corporate Social Responsability concept: environmental, social and corporate governance.

We begin by explaining the fallacies of the linear modeling common to most of the studies in the area, fallacies related to variable selection and to model validation. We then relax the assumption of a linear relationship between the CSR scores and profitability and conduct a non-linear analysis based on performant approaches from the statistical learning literature (smooth spline enforced by boosting and regression trees enhanced through bagging). We find that non-linear methods bring important improvements in explaining profitability and so demonstrate that an increased level of statistical sophistication is needed if one wants to uncover subtle relationships explaining firm performance.

A number of Corporate Social Responsability variables like Corporate Governance, Codes of Conduct and Human Capital seem to have important explanatory power. In particular, the Corporate Governance score positions ahead of well-known variables that explain profitability as leverage, capital intensity, change in sales or size. Above average SAM scores in Corporate Governance can add as much as 3 to 4% to firms's Return on Assets (depending on the method used).

However, most of the CSR scores considered in the analysis do not seem to contribute to explaining profitability. In particular, the scores related to environmental dimension are not among the significant variables. Moreover, the Corporate Social Responsability variables account for little of the improvement in prediction error (most of the gains reported are due to taking into account the non-linear nature of the relationship between the explanatory variables and the measure of profitability). They improve the prediction accuracy with 1 to 3% (depending on the method employed).

The results suggest that company managers do not face a tradeoff between eco-efficiency and financial performance, and that investors can use environmental information for investment decisions.

[5]SAM group in cooperation with Dow Jones Indexes and STOXX Limited publishes the Dow Jones Sustainability Indexes (DJSI).

1. Introduction

Over the last decade the financial community has paid increasing attention to social responsible investments and to their ability to influence the social responsible behavior of corporations. The extent of interest is such that nearly 9.5% of the assets under professional management in the US in 2005 (according to Social Investment Forum, 2005) are selected on sustainability related criteria. The general feeling in the written media[6], and not only, is that making investments on socially responsible criteria is gradually becoming a mainstream interest rather than a niche one.

In its stronger form, the concept of Corporate Social Responsibility (CSR) asserts that corporations have an obligation to consider the interests of customers, employees, shareholders, communities, as well as the ecological "footprint" in all aspects of their operations. This obligation is seen to extend beyond their statutory obligation to comply with legislation. Initially understood as pure ethical or environmental concern, Corporate Social Responsability has developed into a complex concept spanning on three main directions: environmental, social and governance (United Nations Environment Programme Report 2007).

The concept of Corporate Social Responsability is closely linked with the principle of Sustainable Development, which argues that corporations should make decisions based not only on financial factors such as profits or dividends, but also based on the immediate and long-term social and environmental consequences of their activities. The concept of Sustainable Development, first introduced by the World Commission on Environment and Development (1987 Report, pg. 8), was defined as meeting "the [human] needs of the present without compromising the ability of future generations to meet their own needs".

The debate on the role of corporations in society and on the extent to which their social responsible behavior can be responsible for superior financial performance can be traced back to the beginning of the 1970s (Moskowitz 1972, Vance 1975)[7]. Meanwhile, the perceived role of corporations, actors on a global scene, has become even more important while the willingness of investors to ensure compliance with their beliefs has grown even stronger. At present, firms are seen not only as responsible to their direct owners, i.e. shareholders and debt holders, but to stakeholders as a whole, an enormous step-ahead from the era when corporations were profit-maximization entities only.

[6] *e.g. "Public pensions leading the way on SRI", Financial Times weekly review of the fund management industry, May 21,2007*

[7] Socially responsible-type of investments based on the "self-referential screening" (i.e. excluding companies that one doesn't want to own from the investment universe) has existed for more than 70 years (Dillenburg et al. 2003). The range of exclusionary screens, already in place in the mid 1920s when it covered tobacco and alcohol, extended gradually to include by the end of the 1990s gambling, production of handguns, racial discrimination and human rights issues (Kinder and Domini 1998). The last decade have seen a further extension of the content of such screens.

The increased interest in environmental, social and governance issues stimulated a dynamic development of econometric and financial literature focusing on the relationship between corporate social performance and firm profitability. Despite the boom in the number of papers on the issue during the last decade, clear cut answers are still to be provided.

The main difficulty (as well as one of the reasons for obtaining conflicting results) consists in defining adequate and representative *quantitative* measures for the complex Corporate Social Responsability concept. One cannot emphasize enough that the results of any analysis set to evaluate the impact on firm profitability of compliance with the principles of Corporate Social Responsibility depend *essentially* on how well the quantitative measures used in the analysis capture the operational implementation of the multi-faceted, often subjective concepts they claim to measure. In other words, such studies succeed in investigating the association between Corporate Social Responsability standing and firm profitability only to the extend to which the scores they use are relevant measures of the operational implementation of the Corporate Social Responsability principles. Hence the capital importance of the Corporate Social Responsability measures involved and the need for a critical attitude towards the conclusions of any such study.

Methodological differences have added to the variability of results producing not rarely contradicting conclusions. Initially, a large body of literature compared the performance of portfolios (already made ones - the case of mutual funds studies or constructed by the researcher) of stocks chosen based on CSR criteria with that of conventional portfolios (Statman 2000, Derwall 2005 among others). Even when accounting for the lack of homogeneity of the mutual funds in applying screening criteria (Bauer et al. 2005), the portfolio studies provide no clear evidence of superior stock performance on the side of the "best-in-class" CSR performers. Besides, one serious drawback of such studies is that the CSR concept is often limited to the environmental dimension (see Derwall 2005 and the references therein).

A more recent methodological framework (to which our paper belongs) emphasizes a multi-dimensional approach to the Corporate Social Responsability concept. The studies in this branch of literature use multi-dimensional CSR firm scores provided by independent ranking companies to conduct an analysis at company-level. Although the environmental dimension is still predominant (see King, Lenox 2001, Guenster et al. 2005 among others), other CSR criteria (such as employees and community activities) have been recently taken into consideration (Brammer et al. 2005). A possible critique to most of these studies refers to the small (less than four) number of criteria to quantify the Corporate Social Responsability concept.

As a response to such a critique, our analysis is conducted on a fourteen-dimensional firm specific data set covering all three environmental, social and governance aspects of the Corporate Social Responsability principles. The data set was provided by Sustainable Asset Management

(SAM) group, an independent asset management company specialising in sustainability investments[8].

The main goal of our analysis is assessing the relevance of corporate, social and environmental performance variables to explaining firms operational performance as measured by its return on assets[9]. We chose the return on assets as a dependent variable primarily because it is one of the broadest measures of firm operating performance[10]. We pay close attention to potentially confounding influences by including a broad range of control variables. As a by-product of the analysis we also evaluate the relative contribution of accounting and market information based variables that are usually used in profitability studies.

Our paper is set to give a methodological example of how to evaluate the relevance of a *given set* of Corporate Social Responsability measures to explaining firm profitability. The analysis is performed under the caveat that the findings of any such an analysis shed light on the relationship between compliance with Corporate Social Responsability principles and profitability *only to the extent* to which the scores at hand succeed in quantifying the concepts they set to capture. The emphasis of our study is on an accurate statistical analysis, based on the most recent developments in statistical learning theory and practice.

The contribution of our paper to the literature is two-fold. First, we try to address the multi-faceted nature of the Corporate Social Responsability concept by using a high dimensional set of scores covering all of its three main aspects: environmental, social and governance. We believe that a successful investigation of the relation between corporate, social and environmental performance and firm profitability is possible only if based on scores that try to quantify the complex nature of the Corporate Social Responsability concept.

Second contribution is methodological in nature. We argue that the methodology of linear regression commonly applied in studies like ours is affected by a number of fallacies (the most serious ones being the choice of the variables to be included in the model and the over-simplification of the relationship to be described through a rigid restriction to linearity). We advocate a more refined approach based on the results from the modern field of statistical learning.

---

[8]SAM, in cooperation with Dow Jones Indexes and STOXX Limited, publishes and licenses the Dow Jones Sustainability World Indexes (DJSI), a series of global sustainability benchmarks launched in September 1999. The indexes are based on SAM's corporate sustainability assessment, which identifies global sustainability leaders on the basis of economic, environmental and social criteria.

[9]The impact of CSR performance variables on financial profitability as measures by investors/market firm valuation is left for further research.

[10]Given the 'case study' flavour that we wanted to give to this article, i.e. an in-depth examination of a single instance: a case, we limited our interest to this accounting-based measure. The methodological approach that we advocate can, of course, be used with other performance measures. These analyses will be performed elsewhere.

2. Theories about the link between Corporate Social Performance and Firm Profitability

To our knowledge, no consensus have been reached in the economic or financial theory relative to the link between corporate social performance and firm profitability. Although in this article we are mainly interested in firm's profitability from an *operational performance* point of view (the profitability measure we employ is firm's return on assets) for the sake of completeness we briefly review also the theories as well as the empirical evidence to date behind a possible connection between corporate social performance and market-based financial profitability.

2.1. **Economic theory.** Economic theory related to corporate social responsibility is mainly based on the environmental component[11]. On one hand, there are the views expressed in the traditional neoclassical economics (Cohen et al. 1995) based on increased costs argument. Stringent environment regulation is believed to lead to higher compliance costs for companies within sensitive sectors of the economy (eg. oil and gas), implying a competitive disadvantage. As the environment is one of the four main production factors, imposing limitations on it (through investments in cutting edge technologies as a way to reduce pollution, for example) will necessarily increase the costs (Palmer 1995, Siebert 1980).

On the other hand, the revisionist economic view proponents (Porter, van der Linde 1995 among others) base their theory on cost savings and revenue increases argument. Stringent Corporate Social Responsability regulation is assumed to generate a competitive advantage when it is complied with through innovative technology. The need of complying with environmental regulation will stimulate new solutions in terms of new technology that improve resource productivity, production process efficiency and avoidance of waste. Also, better technology implies lower environmental risks as less environmental taxes or charges will be paid as well as fewer pollution rights will be purchased (Schaltegger and Muller 1998).

2.2. **Financial theory.** Financial theories on the connection between corporate social performance and firm financial profitability are based on equilibrium asset pricing models as well as on the efficient market hypothesis (Guenster et al 2005 and the references therein). It predicts three possible relations.

One direction of reasoning postulates a neutral relation. It assumes that the risk associated with compliance with Corporate Social Responsibility is not priced, therefore all companies, CSR complying as well as non-CSR complying, have the same rate of expected return and face the same cost of equity capital (Hamilton et al. 1993). This reasoning is in line with standard financial theory (risk-return paradigm) where only risk factors are priced in the market.

On the other hand, if the risk associated to Corporate Social Responsibility compliance is (correctly) priced by the market, the same risk-return paradigm would imply a negative relation between corporate social performance and financial performance. As put forward by Shane and

[11] This is possibly due to the "old" governmental interest for policy implications, i.e. environmental regulation.

Spicer (1983), firms which actively account for the CSR risk factor are seen as less risky investments (relative to the firms that ignore it). Consequently, on a risk-adjusted basis, their expected returns are predicted to be lower.

Finally, the third view postulates that the compliance with Corporate Social Responsability principles is not efficiently priced by market participants. A positive (negative) relation follows depending on the sign of the inefficiency. For example, Hamilton et al. (1993) argue that, if a sufficiently large number of investors underestimate (overestimate) the probability that adverse events related to Corporate Social Responsability issues might affect companies not complying with the CSR principles, then their stocks will provide lower (higher) risk-adjusted return than socially responsible companies stocks.

Since the answer to the question whether the risk associated to Corporate Social Responsibility issues is (correctly) priced by the market cannot be given on theoretical grounds only, it is investors perception of the relevance of the Corporate Social Responsability principles that counts in the end. If investors believed that companies implementing the Corporate Social Responsability principles are resource wasteful, they would determine a negative return premium on these companies stocks. To the contrary, if CSR behavior of companies is in line with investors beliefs, they would determine a positive return premium for these companies stocks (Ullman, 1985). We turn now towards the empirical evidence. Anticipating, we can say that empirical results have failed so far to capture investors beliefs.

3. Empirical evidence to date

Although our focus is on accounting-based measures of operational performance, we review also prior research that focuses on other financial variables as stock returns or firm value measures as Tobins q. The review is structured according to the aggregation level at which the analysis is performed: first methodologies involving aggregated data (like portfolios of stocks where data on a large number of firms gets aggregated before conducting the analysis) are discussed followed by studies where disaggregated, firm specific information is used.

Anticipating, the empirical results on the relationship between sustainability performance and financial performance seem inconclusive. This lack of consensus in the empirical evidence is fueled by contrasting theoretical models on one hand as well as by differences in methodology employed and choice of corporate social or stock market performance measures used across studies, on the other hand.

3.1. **Mutual funds and portfolio analyses.** A large amount of empirical work on the relation between sustainability and financial performance measured by returns is based on CAPM, APT and Fama French or Carhart multi-factor model and focuses on testing whether there is a significant difference in risk adjusted returns of socially responsible (SR) screened funds/portfolios and conventional funds/portfolios (see Hamilton et al. 1993). Though broadly accepted and recognized, the screening criteria are highly dependent on principles specific to each fund. The analyses performed at mutual fund or unit trust level generally find no statistically significant difference in risk-adjusted returns between the (SR) and non-SR investments. These results, initially obtained through simple one-factor CAPM model (Hamilton et. al 1993, Statman 2000), are further enforced by similar evidence from enhanced performance attribution models like APT or Carhart models (DiBartolomeo and Kurtz 1999, Guerard 1997, Luther and Matatko 1994). The most comprehensive study in this area in terms of sample size, market coverage and methodology is Bauer et al. (2005). The results obtained by using four-factor Carhart model do support the overall conclusion of no extra risk-adjusted returns of SR funds relative to non-SR funds.

Although homogeneous in message, these studies suffer from the criticism of lack of transparency on the screening criteria used by SR funds that most often leads to little or no difference in their composition relative to conventional funds. To address this issue, a more recent strand of literature focuses on building homogeneous portfolios based on sustainability scores provided by different ranking agencies (like ICCR, KLD Research & Analytics, Innovest)[12].

The sustainability dimension that received most if not all attention in this strand of literature is the environmental one. Generally, the methodology consist of our-factor Carhart model together with some linearity assumption. Though mixed results are found with respect to whether there is a premium or a penalty for investing in environmental leaders stocks, evidence for a return premium prevails. In line with Diltz (1995), but contrary to Cohen et al. (1997), Derwall et al. (2005) find a 6% risk adjusted premium for investing in the eco-efficient high-ranked portfolio, after controlling for investment style and industry.

3.2. **Company-level analyses.** An alternative approach to analyzing the relationship between CSR performance and firm profitability is to explore the nature of their association in a regression model, based on the increasing (decreasing) efficiency argument (see section 2.1). Most studies assume that the financial performance measures and firm sustainability scores are related in a *linear* fashion. Most often the firm sustainabiltity measure is uni-dimensional (usually a CSR score or an environmental score). The analysis is performed at firm-level, with different firm profitability measures being employed, while firm and industry specific characteristics are captured through a number of control variables.

The studies focusing on the environmental dimension of the CSR performance provide mixed results; however, evidence for a positive impact of environmental performance measures on firm performance measures prevails.

Hart and Ahuja (1996) find evidence of a positive effect of environmental performance measures on one and two period ahead return on assets and return on equity respectively, while no correlation within the same time period is found. Russo and Fouts (1997) suggesting that environmental performance is positively associated with ROA and that this association is more pronounced for high-growth industries. King and Lennox (2002) suggest that pollution prevention is associated

---

[12] *This approach is, of course, not free of criticism as the ratings themselves might be taunted by subjectivity.*

to higher return on assets while Wagner et al. (2002) document an uniformly negative relation between environmental and financial performance for companies within the pulp and paper industry (through a simultaneous equations framework). Evidence for a time-varying (negative followed by a positive increasing) premium for "environmental winners" is found, through a cross-sectional data analysis of the relation between Innovest environmental performance measures and both Tobin's q and ROA (Guenster et al. 2005).

There is little evidence on the relation between other CSR dimensions and firm profitability. Moreover, there are very few studies where the impact of multivariate measures of Corporate Social Responsability on profitability is evaluated. Brammer et al. (2006), in one of the few studies touching upon other CSR measures than environmental ones, show that stock returns are negatively correlated to employment and environment indicators while a weak positive association is found with a community indicator. The standard performance attribution methodology is employed together with the above-mentioned CSR measures as explanatory variables.

4. Corporate Social Performance and Firm Profitability measures

4.1. **Corporate Social Performance measures.** Understanding the relationship between corporate social and financial performance is made difficult by the lack of consensus on exactly how to transform into numbers the principles of Corporate Social Responsibility. Various proxies have been used over the years as measures of corporate social performance with an evolution that goes from highly subjective (such as surveys of business students, see Heinze 1976, or business faculty members, Moskowitz 1972, or the Fortune rankings , see McGuire et al. 1988, Herremans 1993 or Preston and OBannon 1997) or extremely specific (official corporate disclosures, annual reports to shareholders, CSR reports as in Hamilton 1995, Hart and Ahuja 1996, Karpoff et al. 1998, Moughalu et al. 1990) to more standardized and comprehensive measures. These are sustainability scores or rankings that represent an aggregate measure of compliance with the principles of Corporate Social Responsibility and are provided by specialized, independent agencies such as KLD Research & Analytics, Innovest, Sustainable Asset Management Group, etc. The scores are constructed through in-house research based on corporate public and private documentation.

Despite the high level of standardization and the special care applied in the process of constructing the scores, it is almost impossible to determine whether these more sophisticated social performance measures are objective since the original source of information rests with the company itself and few companies have their Corporate Social Responsability reports externally verified. Thus, corporate social performance measures are still subject to questions about impression management and subjective bias.

We approached the choice of a relevant set of scores for our analysis with this caveat in mind.

4.1.1. *SAM Group Corporate Social Responsability measures.* By choosing the Sustinable Asset Management Group, one of the leading institutions specializing in sustainability investments, as our data provider we tried to capitalize on SAM's many years of sustainability expertise. The relevance of its data set is supported by the fact that SAM group, in cooperation with Dow Jones Indexes and STOXX Limited publishes and licenses the Dow Jones Sustainability World Indexes (DJSI), a family of indexes that provides a growing number of asset managers with objective and professional benchmarks for sustainability investments. The Dow Jones Sustainability Indexes (DJSI) were the first global sustainability indexes to be launched (in 1999). Currently the family consists of major market benchmarks as Global, European, Euro zone, North American and US, etc. Since their inception, the total volume of financial products based on this family of indexes and issued by over sixty financial institutions in fourteen countries totals more than 5 billion USD. Hence, the prominence of the data provider on the scene of sustainable investment was an important stimulus for our decision on the CSR measure to use.

Another important reason for our choice was the comprehensive, multi-dimensional nature of the CSR measure built by SAM Group. It consist of fourteen scores reflecting firm compliance with the principles of Corporate Social Responsability . The scores cover all three main CSR dimensions: economic, social and environmental.

The economic dimension is closely related to the ethics of conducting the business, i.e. to how the company manages itself and the relations with third parties. There are five criteria/scores quantifying the relevant information in this dimension. They are Corporate Governance, Investors Relations, Risks Crises Management, Codes of Conduct/Corruption/Bribery, Customer Relationship Management.

The group of social criteria reflects the values and practices of the company with respect to those directly involved in the profit maximization process, such as employees, shareholders. There are five relevant criteria/scores belonging to this group: Labor Practice Indicators, Human Capital Development, Talent Attraction/Retention, Stakeholder Engagement, Corporate Citizenship/Philanthropy.

The environmental dimension of CSR is being tackled from two different angles. One of them reflects the importance paid on environmental issues, the extent to which the company has an environmental policy. The other one captures the environmental performance itself defined in terms of efficient use of resources and greenhouse gas emissions reduction. The criteria/scores are Environmental Policy and Eco-efficiency, respectively. Two additional criteria/scores, Social Reporting and Environmental Reporting, are also relevant to this dimension. Their special standing is due to the way they are built. While all the other scores benefit mainly from private information provided by the company, the last to scores are constructed based on publicly available information: CSR annual report, CSR part of the annul report, company website or other references to the CSR behavior of the company.

The major source of information used in constructing the fourteen scores are the answers to SAM Questionnaire that the firms are requested to fill in[13]. The SAM Questionnaire is been improved every year so that it incorporates new aspects that become topical in corporate social

[13]For details see http://www.sustainability-indexes.com/06_htmle/assessment/overview.html

responsibility. The process of obtaining company specific scores takes place in several steps. First, the companies asked to participate in the annual review (these companies represent at least 20% of the market capitalization in all industries and in all the major economies of the world) fill in a general questionnaire covering the three relevant areas of Environment, Social and Governance. The companies need to support the information disclosed with appropriate documentation.

Next, the answers for all questions are converted into grades. Two averaging processes follow. In the first one the scores for each of the fourteen criterion described above are obtained (this are the scores that we analyze in the sequel). A final company specific score describing its overall CSR standing is also produced by averaging the fourteen grades obtained in the previous step. This final score is then used to identify sustainability leaders to be then included in the Dow Jones Sustainability Indexes. Industry-specific weights are used in the process of aggregation. For example, to obtain the final score, the grades corresponding to environmental criteria will be differently weighted for a company in the financial sector relative to one in the chemical industry sector. Moreover, once the final scores are computed, a telephonic conference is established between a company representative and the SAM financial analyst in order to discuss the main findings.

To conclude, a number of precautions have been taken by SAM Group to limit the impact of the above mentioned subjectivity critique. First, as already mentioned, the companies need to support the information disclosed with appropriate documentation. Then, a special effort has been made to limit qualitative answers through a system of predefined multiple-choice questions and a clear set of rules that convert the answers in numbers of points. Moreover, regular checks are run to verify the truthfulness of the questionnaire answers. Finally, the process of constructing the scores is highly standardized and externally assessed by Price Waterhouse Coopers company. All these measures can alleviate to a certain extent our preoccupation with the subjective bias without fully removing it.

4.2. **Firm Profitability measures.** If an objective quantitative measure of firm compliance with the Corporate Social Responsability principles seems illusive at present, profitability seems easier to evaluate. Based on financial theories, a large number of studies use market returns (on individual stocks, portfolios or mutual funds) as a measure of financial performance (McGuire et al. 1988, Bauer et al. 2005, Derwall et al. 2004). As a good indicator of intangible value (Dowell et al 2000), Tobin's q is considered to be a good proxy for firm value (Guenster et al. 2005). The interest of the studies focusing on market-based performance measures is to provide support for one of the three hypotheses described in section 2.2.

When research studies focus on investigating the impact of compliance with Corporate Social Responsability principles on corporate profitability and efficiency, accounting-based operating performance measures such as Return on Assets (ROA), Return on Equity (ROE) or Return on Sales (ROS) (see King, Lenox 2001 and the references therein) are employed. The hypothesis to be tested is that incorporating CSR principles in the business activities leads to improved (decreased) operational efficiency yielding higher (lower) profitability ratios, as described in section 2.1 .

4.2.1. *Our profitability measure.* The goal of the paper is to investigate if and how compliance with Corporate Social Responsability principles influence the overall firm efficiency and profitability in the ways suggested by the economic theory described in sub-section 2.1. Note that the theoretical discourse there mentioned, although focused on environment can be extended to other CSR dimensions as well. The additional costs incurred by the firm when complying with Corporate Social Responsability principles might (or might not be) offset by higher (lower) revenues due a more innovative technology (for the environment dimension), new and more training programmes, more flexible promotion alternatives (for the social dimension), increased governance transparency (for the governance dimension), etc. Towards this end, we have chosen as profitability measure the Return on Total Assets (ROA) defined as the ratio between net income and total assets. This choice is motivated by the fact that ROA is one of the broadest measures of firm operating performance (Russo and Fouts 1997, Guenster et al. 2005).

4.3. **Variables that Explain Firm Profitability.** Firm profitability is a central issue in economics. A number of variables have been shown in the literature to help explain expected profitability. At the same time, a relation between most of these variables and corporate social performance has been proved in the related literature; therefore they must be included in any analysis on the impact of adopting of Corporate Social Responsability principles. We will briefly discuss the variables commonly encountered in the studies on firm Return on Assets. Note that our analysis, besides its focus on the impact of CSR standing on profitability, will yield conclusions on the nature and the strength of the association between these variables and firm operational profitability as measured by ROA. To our knowledge, our study is among the first to allow for a *non-linear* relationship. As we will see in Section 8, some of the associations are strongly non-linear and accounting for their true nature produces significant improvements in explaining ROA.

Since firms target dividends to the permanent component of earnings, dividends are claimed to have information about expected earnings (Miller and Modigliani 1961). Moreover, Fama and French (1999) find that firms that do not pay dividends tend to be much less profitable than dividend payers. An explanatory variable accounting for these findings is the ratio of dividend to book value of common equity.

Since the market value of a firm is the current value of future net cash flows, measures of market valuation, like the ratio of market value to book value of common equity, are commonly used is studies on expected profitability. It is assumed that market valuation measures pick up variation in the expected profitability missed out by dividends (Fama and French 2000).

Previous research (Ullman 1985, McWilliams, A., and D. Siegel 2000) showed that industry affects a firms performance through industry-specific factors such as competitive intensity and economies of scale. Also, the importance of different CSR criteria is not similar across industries, i.e. the environmental dimension is likely to have a much higher impact on profitability ratios of firms in the heavy industry, relative to firms in the financial industry.

Together with industry, the size variable is used as an explanatory variable in most studies of performance (usually in the form of the log of market capitalization). It is hypothesized to be weakly negatively correlated with performance. Size has been found (Ullman 1985, McWilliams, A., and D. Siegel 2000) to affect not only a firms performance but also the CSR performance. Larger firms seem to adopt the CSR principles more often due to the fact that they can afford allocating more resources to the adoption of CSR standards. According to Burke et al. (1986) as they grow, firms attract more attention from stakeholders and hence are more under pressure to obey to the Corporate Social Responsability principles.

Measures of capital intensity like ratio of depreciation expense to total assets or capital expenditure divided by total property, plant and equipment are also claimed to explain performance (Fama and French 2000). A possible explanation could be that higher capital intensity should be associated with higher performance because investment in technical capital results in knowledge enhancement leading to product and process innovation. Ultimately, innovation leads to enhanced efficiency (McWilliams and Siegel 2000).

Leverage is believed to be a potentially important variable related also to the governance structure and ownership. A negative relationship is hypothesized between leverage and the performance of the company reasoned by an argument saying that firms with strong financial performance prefer not to borrow. Leverage is possibly negatively correlated to a firm CSR standards as well, in the sense that highly leveraged firms are less able to make long-term investments necessary to enhance CSR performance (Dowell et al. 2000).

Retained earnings is a measure of the capacity of the firm to reinvest in its core business. In most cases, companies retain their earnings in order to invest them into areas where the company can create growth opportunities, such as buying new machinery or spending the money on more research and development or to pay off debt. A positive relationship between profitability and past retained earnings is hypothesized. Positive correlation is assumed between retained earnings and CSP, as firms with higher retained earnings have readily available resources for projects leading to CSP improvement.

Although McWilliams, A., and D. Siegel (2000) suggest that R&D costs help explain firm performance as well as CS performance, we follow Waddock, S. A., and Samuel B. Graves (1997) and take these differences into account indirectly controlling for industry.

Finally, a positive association between firm growth and performance has been extensively documented (see Capon et al. 1990 among others). Usually, the proxy used for firm growth is the average percentage change in sales or assets.

5. Methodological Motivation

5.1. **Why model selection and model validation are important.** Most of the quantitative studies on the impact of adopting Corporate Social Responsability practices expend a considerable effort in assessing the association between measures of corporate social performance (CSP) and measures of firm profitability. Determining such a relationship is motivated by two goals: to gain explanatory insight into how the economy operates and to gain predictive power to forecast firms' future profitability. The success of any method trying to establish such an association depends hence on the ability to separate between *systematic* and *stochastic* components. The resulting model should not capture any data set specific details, i.e. noise contribution. This goal is the central premise behind the two important phases of fitting a model: *model selection* (allowing only the significant variables in the model) and *model validation* (evaluation of model predictive performance).

Model selection means selecting a subset of predictor variables that explain a statistically significant proportion of variation in the dependent variable. The capacity of a model to discriminate between systematic and stochastic components is measured in the validation step. A model has little merit if its predictions are not validated. Moreover, the validation step offers the possibility to compare the predicting capabilities of different models.

The traditional approach to validation has been to use the same data both to construct the model and to evaluate its predictive performance. This approach is known to bias the valuation accrediting the model with better performance than its true one. The model, optimized to deal with the particular structure of the data set on which is the estimation is performed, will have less predictive power on another similar data set (Mostlerand Tuckey 1977, Efron 1986, Chattfield 1995). For a more realistic estimate of prediction error, the model should be validated on data that is independent of the data on which the model is built, i.e. the relevant variables are selected and the parameters of the model estimated. Cross-validation, a technique proposed originally for this purpose (Allen 1974, Stone 1974, Geisser 1975), splits the data in two groups of observations, with one group used to model building (the so-called training sample) and the other group reserved for model validation (the so-called test sample).

5.2. **Model selection/validation for multiple linear regression.** By far the most commonly employed statistical technique for establishing the mentioned association is multiple linear regression analysis based on the assumption that there exists a linear relationship between corporate social performance and profitability, obscured to varying degrees by other factors, including noise. However, despite its simplicity, the linear model as it is commonly used in practice, is plagued by a number of problems both in the model selection as well as in the model validation stage.

One well-known drawback of the linear regression model is its potential instability when the hypothesis of linearity is not realistic. That means that when the relationship in the data is not well described by the linear assumption, removal or addition of a few observations can dramatically change the set of optimal variables as well as the estimated coefficients. This weakness can be strongly alleviated by applying the boosting methodology that we describe in the sequel.

Various model selection techniques have been developed for the linear regression (including forward selection, backward elimination, stepwise and exhaustive search). All these procedure involve repeated tests of significance in their search of the optimal set of explanatory variables, a method long recognized by statisticians to result in erroneous inclusion of variables in the final model (Hocking 1976, Miller 1984, 1990). It is the sequential testing of a large number of related models in search of the optimal one that increases the likelihood of including in the final model of predictor variables that have a true random association with the dependent variable. This bias in model selection methods has stimulated research in the statistical field leading to introducing alternative, improved techniques, such as bootstrapping for model selection (Friedman and Peters 1984, Breiman and Spector 1992, Shao 1996). Some of the analysis that follows will address these issues.

Even more problematic than methodological issues related to the choice of the model and its explanatory capacity is the strong assumption of linearity made by the existing studies. There is hardly any reason (hence no *economic* reason) besides that of convenience (easy estimation, apparently easy interpretation of the results) to suppose linear relations between the CSP variables and performance. It is in fact very likely that financial performance dependency on CSP is strong non-linear. It is intuitively pertinent that investing the same amount in becoming more eco-efficient when one is at the frontier of best practice is likely to yield less than investing the same amount when one is a laggard in the field. Our analysis relaxes the assumption of linearity and focuses on a number of non-linear modelling approaches that have emerged in the last decade in the field of statistical machine learning.

5.3. **A Statistical Learning view on establishing associations.** In this paper we bring to attention the bagging and boosting prediction techniques. These methods were developed as an alternative to linear models and avoid the problems that plague them (enumerated above). They over-fit less (or do not over-fit at all), they allow for non-linearities in the relationship between the explanatory and the independent variables and, most important of all, they explain the relationships in the data significantly better.

Both of these methods operate by estimating a model (that can be unstable - as the linear model discussed above is) many times on different sets of observations (so-called training sets) obtained from the original data set. In bagging (Breiman 1996) each training set is constructed by re-sampling the original data.

Boosting (Freund and Schapire 1996) applies a set of weights over the original data set (when re-sampling) to produce the training sets and adjusts these weights after each estimation of the model of the current training-sample. The adjustments increase the weight of examples that are not well-explained by the last-estimated model and decrease the weight of examples that are well explained by the estimated relationship. The final model is derived by simple aggregation.

Bagging (and to a lesser extend boosting) can be viewed as ways of exploiting the this instability to improve explanatory accuracy.

5.4. **Our contribution.** We first use the commonly employed traditional methods (as the step-wise search) for model selection in multiple linear regression analysis. Next, we assess the performance of theoretically sound variable selection methods for linear regression (more precisely, Breiman's bootstrap approach and boosting component-wise linear bases approach; see sub-section 6.3 for details). In the context of linear modelling this approaches are known to produce a more truthful set of explanatory variables.

Next we switch to non-linear modelling and evaluate the ability of different model-selection methodologies proposed in the statistical learning literature to produce models that explain better the data, i.e. that have smaller validation error. We evaluate, discuss and interpret these models.

6. Statistical Learning techniques

We assume that the data are realizations of random variables $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from a stationary process (not necessarily i.i.d.) with p-dimensional predictor variables $\mathbf{X}_i$ and one dimensional response variables Y_i. Examples of Y include the Return on Assets (ROA), the Return on Equity (ROE) or the market return. The explanatory variables include the SAM scores and other variables known to explain firm profitability (see the discussion in Section 4.3).

From a statistical point of view the model is

$$Y_i = f(\mathbf{X}_i) + \epsilon_i, \quad i = 1, \ldots, n. \tag{6.1}$$

where $f : \mathbf{R}^p \to \mathbf{R}$ is *possibly non-linear*, while $E\epsilon = 0$. Our goal is to, first, figure out the components of $\mathbf{X}$ that are associated with Y and, second, to estimate the multivariate function f. The success of these two operations will guarantee good prediction of the values of Y given the values of the explanatory variables $\mathbf{X}$.

6.1. **Improving Prediction.** Bagging and boosting both originate from the machine learning research community, and are based on the principle of aggregation of simple rules derived from the data. This idea was inspired by Breiman (1996) who found gains in accuracy by combining several base regressors, sequentially estimated from perturbed versions of the original sample.

6.1.1. *Bagging.* Bootstrap aggregation, or bagging, is a technique proposed by Breiman (1996a) to reduce the variance associated with prediction, and thereby improve the prediction process. It can be used with many regression methods (in particular we will be using it with trees). The idea of the method is simple: many bootstrap samples are drawn from the available data, some prediction method is applied to each bootstrap sample, and then the results are averaged to obtain the overall prediction. The variance is reduced due to the averaging. Bagging works best when the base regression procedure that is being bagged is not very stable. That is, when small changes in the sample can often result in significant differences in the predictions obtained using a specified method, bagging can result in a serious reduction in average prediction error. (See Breiman (1996b) for additional information about instability.)

6.1.2. *Boosting.* Boosting, like bagging, can be used to improve the accuracy of regression methods. We will be using it with linear regressors and spline bases. Unlike bagging, which uses a simple averaging of results to obtain an overall prediction, boosting uses a weighted average of results obtained sequentially from applying a prediction method to various samples. The samples used at each step are not all drawn in the same way from the same population. Instead, the incorrectly predicted cases from a given step get an increased probability of showing up in the sample drawn at the next step. The method hence fits mostly the data that has not been well-fitted so far. For a more rigorous description of the procedure see the Appendix.

In various comparisons with bagging, boosting generally, but not always, does better, and leading researchers in the field of machine learning tend to favor boosting. Moreover there is strong empirical evidence that boosting is generally resistant to over-fitting. When stopping early, which is usually needed to avoid over-fitting, boosting often does variable selection. Hence, boosting, besides improving the prediction error, functions also as a variable selection tool.

Finally a remark important for the analysis in Section 8. When predicting Y using OLS regression in a case for which the proper form of the model is linear, all the explanatory variables are included in the sample $\mathbf{X}$ and the sample size is not too small, the assumptions needed for the least squares to work are met and there is little to be gained from bagging or boosting because *the prediction method is pretty stable*.

But in cases where the correct form of the model is complicated and not known, there are a lot of predictor variables, including some that are not related to the response variable, i.e. they are just noise and the sample size is not really large, bagging and boosting may help a lot, if a generally unstable regression method such as trees is used. We will see in the sequel that neither the use of methods that correct the flaws of common model selection approaches for linear regression nor boosting of the linear model provides much performance improvement in the linear framework. On the other hand, we will see that the use of non-linear models strengthened by bagging or boosting produces a significant reduction of prediction error. These two facts corroborated offer strong methodological evidence in favour of the presence of *significant non-linearities* in the relation under study between firm profitability and factors that explain it.

6.2. **Non-linear models.** The non-linear models considered in this paper are smooth splines (whose performance is improved by boosting) and regression trees (performance enhanced by bagging). Splines are more often used in the econometric analysis and hence we will not insist on explaining them (for the procedure of boosting component-wise smooth spline see Bühlmann and Yu, 2003). By contrast, we will spend a paragraph discussing the less known regression trees.

6.2.1. *Regression Trees.* Tree-structured classification and regression are alternative approaches that are not based on assumptions of normality and parametric model statements, as are some older methods such as discriminant analysis and ordinary least squares (OLS) regression. Trees explain variation of a single response variable Y by repeatedly splitting the data into more homogeneous groups, i.e. dividing the space of the explanatory variables X into regions on which the response variable Y takes more similar values.

A regression tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Initially all of the data in training set (the data that is used to determine the structure of the tree) are together in one big box. The algorithm then tries breaking up the data, using every possible binary split on every dimension of the explanatory variables. The algorithm chooses the variable on which to perform the split that partitions the data in two parts such that it minimizes the sum of the squared deviations from the mean in the separate parts. The process continues until each subset in the partition reaches a user-specified minimum size (the subsets of the final partition are called nodes). Once a partition is built the estimated value for f is simply the average of the Y s in each of the partition's set. See Figure 6.2.1

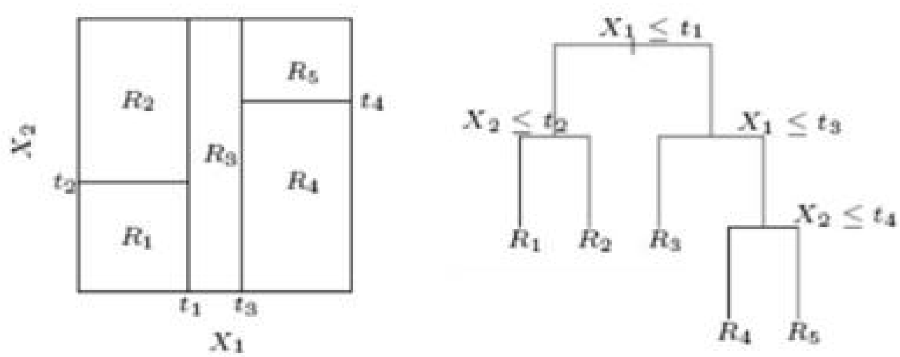

FIGURE 1. Graphical representation of the design space partition together with the graphical representation of the corresponding regression tree.

A regression tree can be represented graphically (see Figure 6.2.1) and this aids exploration and understanding. Trees can be used for interactive exploration and for description and prediction. Advantages of trees include the flexibility to handle a broad range of response types, including numeric, categorical, the invariance to monotonic transformations of the explanatory variables, ease and robustness of construction, ease of interpretation and, last but not least, the ability to handle missing values in both response and explanatory variables.

Regression trees have important advantages with respect to classical linear regression. With traditional regression modeling, one seeks a linear function of the inputs and transformations of the inputs to serve as the response surface for the entire measurement space $\mathbf{X}$. But with a regression tree, some variables can heavily influence the predicted response on some subsets of $\mathbf{X}$, and not be a factor at all on other subsets of $\mathbf{X}$.

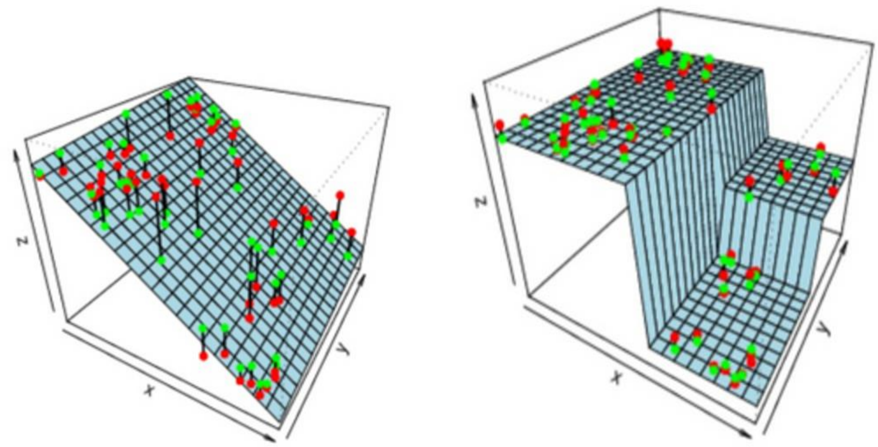

FIGURE 2. Linear regression vs. regression tree estimation.

6.3. **Model Selection.** Given the importance of the model selection step in fitting and evaluating a model we describe separately the methods used in the linear and non-linear framework.

6.3.1. *The linear framework.* The relevant variables are chosen first, using the common step-wise selection procedure and second, through the k-fold cross-validation variable selection method (which we will refer as Breiman's variable selection approach, see Breiman and Spector 1992). As mentioned already, the step-wise variable selection procedure involves repeated tests of significance in the search of the optimal set of explanatory variables, an approach resulting in erroneous inclusion of variables in the final model (Hocking 1976, Miller 1984, 1990). It is the sequential testing of a large number of related models in search of the optimal one that increases the likelihood of including in the final model of predictor variables that have a true random association with the dependent variable. This short-coming is addressed by the so-called k-cross-validation variable selection method to be described in the sequel.

For a given sample, a partitioning in k roughly equally-sized sets is done (therefore, the name of k-fold CV). Repeatedly, one set (to which we refer as the validation sub-sample) is left out for validation purposes whereas the other $k-1$ (which we refer as estimation sub-sample) are used for estimation of a sequence of linear regression models that differ through the number of explanatory variables. An increasing number of variables from one to the total number available is considered. The model is identified by the number of explanatory variables to be considered. For a given model/number of explanatory variables to be included, the selection of variables is made using some well-known selection procedure (such as "best subsets", stepwise forward variable addition or stepwise backward variable deletion). Every such model, i.e. the 'best' model with a given number of explanatory variables, is estimated on the estimation sub-sample and is then used to predict the observations in the validation sub-sample. The prediction error is recorded.

The procedure is repeated for all possible k validation sub-sets. The total prediction error of every model is obtain as the average of the prediction errors of the k instances. The model (identified by the number of explanatory variables to be considered) with the smallest overall prediction error is chosen.

Closely following the results in Breiman and Spector (1992), we have implemented a 5-fold cross-validation[14]. The variable selection method used was backward elimination.

Cross-validation method represents an improvement over classical variable selection due to several reasons. First, over-fitting is avoided as estimation and validation are performed on two different samples. Second, selecting variables based on prediction error on a separate set does not suffer from the selection bias induced by frequently used selection criteria such as C_p, Akaike Information Criterion(AIC) or Bayesian Information Criterion (BIC).

6.3.2. *Non-linear framework.* Boosting is also a variable selection tool. An early stopping rule, usually needed to avoid over-fitting, yields also a method of variable selection.

In the case of bagging, the choice of the relevant explanatory variables will be made based on the partial contribution of the variables as well as on their relative importance (see also the next subsection).

6.4. **Interpretation.** Since $f(\mathbf{x})$ is a multivariate function that depends on a large number of variables, and hence difficult to visualize, interpretation of the estimation results is of utmost importance. This includes gaining an understanding on which are the particular input variables that are most influential in explaining the variation of f as well as on the nature of the dependence of f on those influential inputs. To the extent that estimated f at least qualitatively reflects the nature of the true function f, such tools can provide information concerning the underlying relationship between the explanatory variables $\mathbf{X}$ and the dependent variable Y. In this section, two tools are presented for interpreting tree bagging/boosting approximations. We follow the presentation in Friedman (1999). These interpretative tools are used in the data analysis in Section 8.

6.4.1. *Relative importance of input variables.* The first tool introduced is the relative influences I_j, of the individual inputs X_j, on the variation of $f(\mathbf{x})$ over the joint dependent variable distribution.

$$I_j = \left(E\left[\frac{d\hat{f}(\mathbf{x})}{dx_j}\right]^2 \mathrm{var} X_j \right)^{1/2} \tag{6.2}$$

For a given sample and a regression tree T estimate of f, a sample version $\hat{I}_j(T)$ of the quantity (6.2) can be introduced (see Friedman 1999). For a collection of regression trees $\{T_m\}_{m=1}$, as for example, the result of boosting or bagging, i.e. for a family of regression trees that in a linear

[14] This is based on their finding that leaving out 20% of the data gives a better performance in CV than complete CV.

manner yield an estimate of f, the sample version of I_j is constructed taking a linear combination of the ones corresponding to the individual trees:

$$\hat{I}_j = \frac{1}{M} \sum_{m=1} \hat{I}_j(T_m) \tag{6.3}$$

6.4.2. *Partial dependence function.* Visualization is one of the most powerful interpretational tools. Unfortunately, it is limited to low-dimensional arguments. Viewing functions of higher-dimensional arguments is more difficult. It is therefore useful to be able to view the partial dependence of the approximation $f(\mathbf{x})$ on selected small subsets of the input variables. Although a collection of such plots can seldom provide a comprehensive depiction of the approximation, it can often produce helpful clues, especially when $f(\mathbf{x})$ is dominated by low-order interactions.

Consider the sub-vector $\mathbf{X}_{\mathcal{S}}$ of $l < p$ of the input predictor variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$, indexed by $\mathcal{S} \in \{1, 2, \ldots, p\}$. Let $\mathcal{C}$ be the complement of $\mathcal{S}$. A general function $f(\mathbf{X})$ will depend in general on all of the input variables: $f(\mathbf{X}) = f(\mathbf{X}_{\mathcal{S}}, \mathbf{X}_{\mathcal{C}})$. The *partial dependence function* is defined as the effect of $\mathbf{X}_{\mathcal{S}}$ on f after accounting for the average effects of the other variables $f(\mathbf{X}_{\mathcal{C}})$

$$f_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}) = E_{\mathbf{X}_{\mathcal{C}}} f(\mathbf{X}_{\mathcal{S}}, \mathbf{X}_{\mathcal{C}}). \tag{6.4}$$

A sample version of (6.4) can be defined as

$$f_{\mathcal{S}}(\mathbf{X}_{\mathcal{S}}) = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{X}_{\mathcal{S}}, \mathbf{X}_{i\mathcal{C}}), \tag{6.5}$$

where $\mathbf{X}_{1\mathcal{C}}, \ldots, \mathbf{X}_{N\mathcal{C}}$ are the values of $\mathbf{X}_{\mathcal{C}}$ occurring in the data sample. For additive tree models (like the one produced by boosting or bagging), the partial dependence functions are averaged over the constituent trees.

Variables' relative importance (6.3) together with their partial dependence functions will be used as essential tools for variable selection for the estimation of f by bagging regression trees in the data analysis in Section 8.

The two approaches, i.e. boosting to the component-wise smooth spline and bagged regression trees, were selected as being among the most performing methods that are at the same time interpretable among the many non-linear regression approaches from the statistical learning field. We cannot emphasize enough how different the two methodologies approach the estimation of the associations in the data.

7. Description of the data set

As mentioned the firm profitability measure we consider is Return on Assets (ROA), an accounting financial measure defined as firm earnings/profit after interest and tax per unit currency invested.

Among the explanatory variables we count the SAM scores for fourteen CSR criteria. Each score is obtained by summarizing the answers of up to 11 questionnaire questions, both qualitative and quantitative in nature. The scores range from 0 to 100, with 100 being the highest possible value. An example of a record in the sample is provided in table 1.

Codes of Conduct	68.1
Corporate Citizenship	56.7
Corporate Governance	80
Customer Relationship Mgt.	73
Eco-Efficiency	20
Environmental Policy	81.7
Environmental Reporting	16
Human Capital Development	53
Investor Relations	70
Labor Practice Indicators	84.72
Risk Crisis Management	90.90
Social Reporting	45.5
Stakeholder Engagement	44.8
Talent Attraction/Retention	70.3

Table 1. Example record.

As stated, the central issue in this paper is the attempt to capture the impact of CSR practices might have on firm profitability. Seven other control variables that have been found to help explain profitability, as discussed in Section 4.3, are added to the SAM scores. These variables are size (log of total assets), leverage (ratio of long-term debt divided by total assets), price to book ratio, dividends to book ratio, retained earnings, three year average percentage change in sales (hereinafter called change in sales) and capital intensity (defined as the ratio of capital expenditures to fixed assets). They were obtained from Worldscope Datastream. Moreover, eight industry dummies were also included among the explanatory variables. The nine industries they represent, defined in accordance to the MSCI global industry classification standards (GICS) are: Oil and Gas, Industrials, Consumer Goods, Healthcare, Consumer Services, Telecommunications, Utilities, Technology and Basic Materials. We excluded Financials as they are highly regulated and regulation may produce unusual behavior of profitability (see Fama and French, 2000).The companies in the sample represent all of the major economies in the world with roughly 45% being European, 30% North-American and 15% Japonese.

8. Data analysis

In the analysis that follows the firm profitability is measured by the return on assets (ROA) while the independent variables, i.e. the variables that explain profitability are the ones discussed in Section 7 . The analysis is limited to the financial year 2004[15]. The sample size for 2004 is of 386 firms. The histograms of returns on assets as well as the fourteen CSR variables are shown in Figure 8.

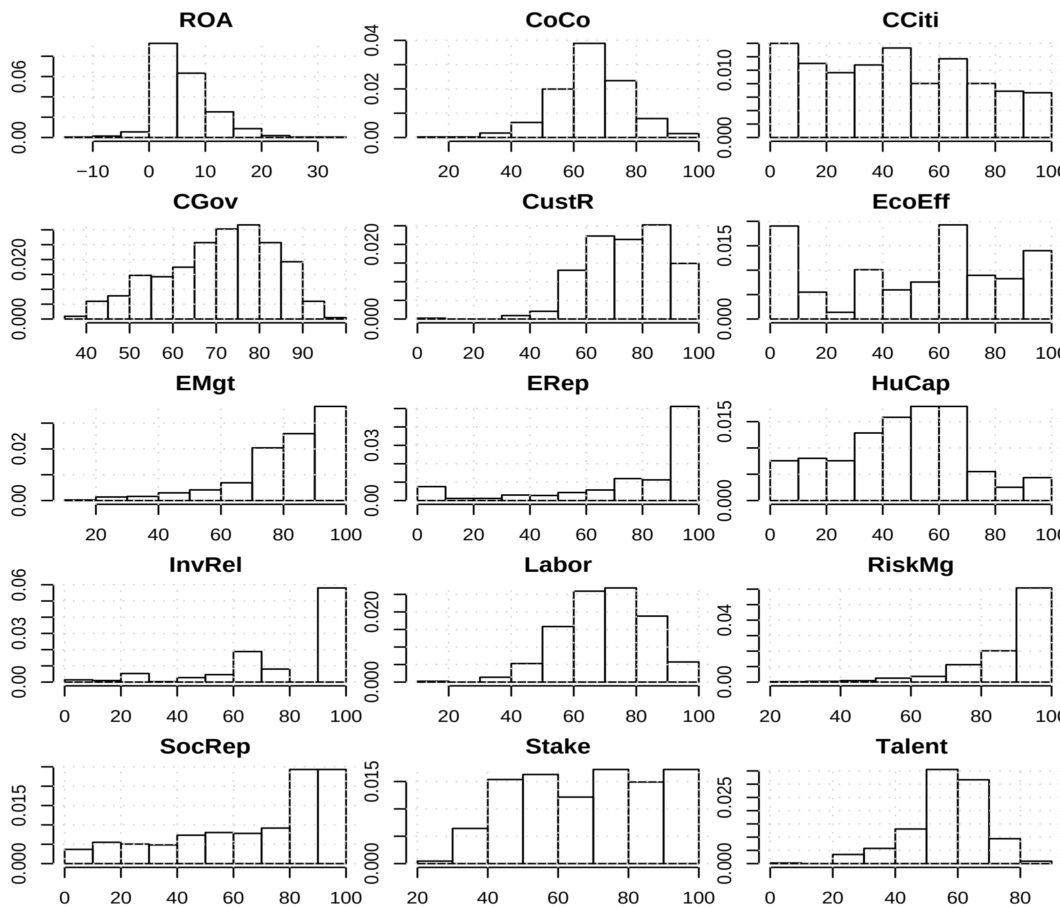

FIGURE 3. Histograms of the independent variable ROA (first graph) and of the fourteen CSR scores.

We note that some of the variables seem less informative than the other. For example, most of the companies got perfect (or almost perfect) scores for Environmental Reporting, Investor Relation Management or Risk Management.

In the regression analysis that follows all the variables are centered. The dependent variables are also standardized such that they have sample variance equal to one. This operation improves the interpretability of the results.

[15]Our analysis focuses on an *in-depths* understanding of the performance of various statistical approaches to evaluation of the possible impact of CSR practices as captured by the SAM scores on firm profitability. For this reason we limit our attention to one profitability measure, i.e. ROA and one financial year, i.e. 2004. Once the methodology is well developed, it can be applied to other measures and financial years.

Several methods for the choice of the significant explanatory variables have been used. Their relative performance has been evaluated through a *cross-validation exercise*, i.e. the models were estimated on a (larger) part of the sample and their explanatory power was evaluated on the remaining (smaller) sub-sample. The procedure was applied on a partition of the sample in five randomly-chosen sub-samples (five-cross-validation). This approach separates estimation from validation yielding a more truthful comparison of the explanatory capabilities of different approaches.

8.1. **Linear framework.** We begin the data analysis assuming a linear relation between firm's profitability as measured by ROA and the explanatory variables discussed in Section 7. (This assumption will be relaxed in the sequel.) We performed a linear regression analysis and made use of three different methods of variable selection: the ubiquitous step-wise algorithm, Breiman's method for variable selection (described in sub-section) and the selection provided by the early stopping rule of boosting needed to avoid over-fitting.

Recall that if the regression method chosen for basis is linear regression the boosting procedure selects at every step the best (not necessarily a different one for each iteration) predictor variable in a simple linear model, i.e. the variable reducing the residual sum of squares most.

Table 2 reports the frequency and the sign of the variables chosen as significant using the three different model selection methods for linear regression: stepwise selection, Breiman's approach and boosting linear bases. Recall that the choice of variable has been performed five times with all methods, on the five different estimation sub-samples created in the five-fold cross-validation exercise. It is worth noticing that the variables' contribution has consistently the same sign (the sign next to the digits in the table indicates the sign of the contribution of the given variable). We also note that variables like price to book, retained earnings, average change in sales, corporate governance are always chosen. Since the performance of the linear model is much poorer than that of the non-linear models, we only report the signs of the coefficients. Including the coefficients would have made the presentation cumbersome without adding much qualitative information.

The performance of the three variable selection approaches is evaluated based on their average prediction error on the validation samples of the cross-validation exercise. The results are presented in the second line of Table 3. The reference model is a linear regression without any variable selection, i.e. including all the variables at hand. There is a wide-spread belief among the practitioners that throwing in all explanatory variables is a good way of dealing with the issue of variable selection. One lets the OLS procedure to chose some coefficient for every variable. As previously discussed, such an approach is bound to produce poor results as confirmed by the numbers in Table 3.

The popular step-wise variable selection rule in the linear regression set-up produces a slight reduction (1.3%) of the prediction error. Applying Breiman's more sophisticated (and theoretically sound) method of variable selection does improve the prediction with almost 4%. In the

Criteria[16]	Stepwise Regression	Breiman's Method	Boosting Linear
Assets ("Assets")	3-	1-	5-
Leverage ("Lev")	4-	3-	5-
Price to Book ("PToB")	5+	5+	5+
Dividends to Book ("DToB")			
Retained Earnings ("REarn")	5+	5+	5+
Average change in sales ("dSales")	5+	5+	5+
Capital intensity ("CEFA")	4+	2+	4+
Codes of Conduct ("CoCo")			1-
Corporate Citizenship ("CCiti")			
Corporate Governance ("CGov")	5+	4+	5+
Customer Relationship ("CustR")			
Eco-Efficiency ("EcoEff")	1-		1-
Environmental Policy ("EMgt")			
Environmental Reporting ("ERep")		1-	4-
Human Capital Dev. ("HuCap")	4+	1+	4+
Investor Relations ("InvRel")	3+	3+	4+
Labor Practice Indicators ("Labor")	3-	1-	5-
Risk Crisis Management ("RiskMg")	4+	3+	5+
Social Reporting ("SocRep")	1-	1-	1-
Stakeholder Engagement ("Stake")	1+	3+	4+
Talent Attraction ("Talent")	3-	1-	4-

TABLE 2. The frequency and sign with which the explanatory variables are included in the model for the five-cross-validation. The number indicates how many times a certain variable has been included in the linear model during the five variable selection operations performed in the cross-validation exercis. The sign next to the digits indicates the sign of the contribution of the given variable to explaining ROA.

linear framework best results are obtained using boosting with component-wise linear base which improves on the regression with all explanatory variables by 4.8%.

As a conclusion, under the rigid assumption of a linear relationship between profitability and the explanatory variables described in Section 4.3, even more sophisticated methods of variable selection, such as Breiman's approach or boosting with component-wise linear bases, produce modest prediction gains. However, as discussed earlier, the assumption of linearity has no real economic base and is simply made for statistical convenience. We will next investigate the relationship profitability/explanatory variables relaxing the assumption of linearity.

8.2. **Non-linear framework.** Two non-linear approaches were considered for modeling the relation of interest: smooth splines and trees.

Generalized linear models, introduced by Hastie and Tibshirani (1986), add more flexibility to the linear structure and have been extremely popular. Their flexibility and precision can be enhanced by boosting (whose framework is especially useful for high dimensional problems). If the regression method is chosen to be smoothing spline the procedure will select at every step the best (not necessarily a different one for each iteration) combination of a variable and a smooth spline function of that variable that reduces the most the residual sum of squares of the model.

Regression trees are nonparametric computationally intensive methods that have greatly increased in popularity during the past dozen years. They can be applied to data sets having both a large number of cases and a large number of variables, and they are extremely resistant to outliers (see Steinberg and Colla, 1995). Regression trees have become widely used among members of the data mining community, although their presence in the econometric literature is scarce.

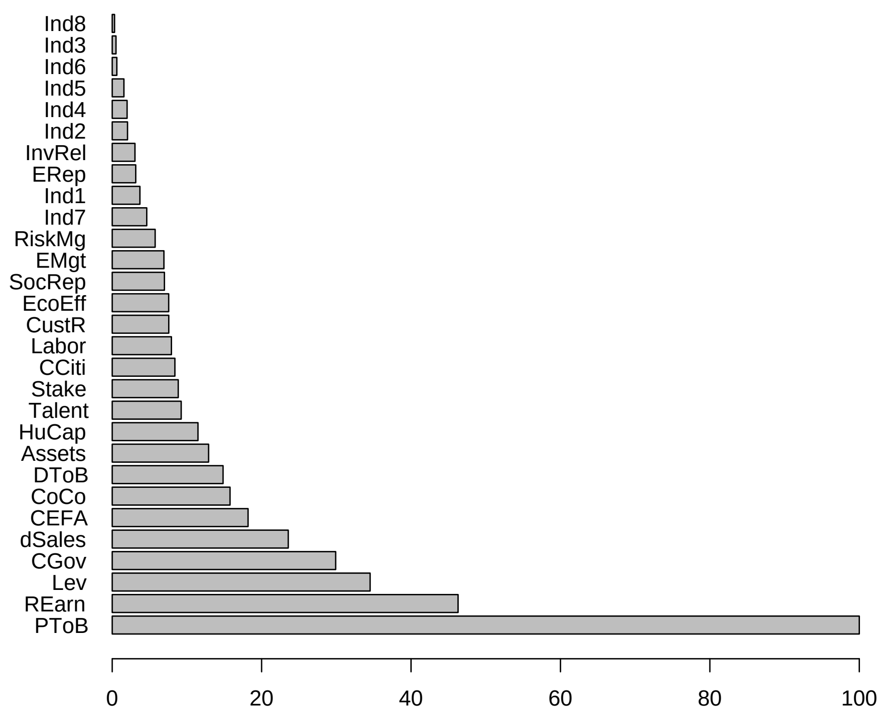

Figure 4. Relative Variable Importance for bagging regression trees.

The performance of both non-linear approaches was improved applying bagging in the case of the tree modeling and boosting to the component-wise smooth spline. In the case of bagged trees the choice of the relevant explanatory variables is made based on the partial contribution of the variables as well as on their relative importance displayed in Figure 8.2 (see also Section 6.4). For the boosting splines the variable selection is a by-product of the procedure itself. Recall that

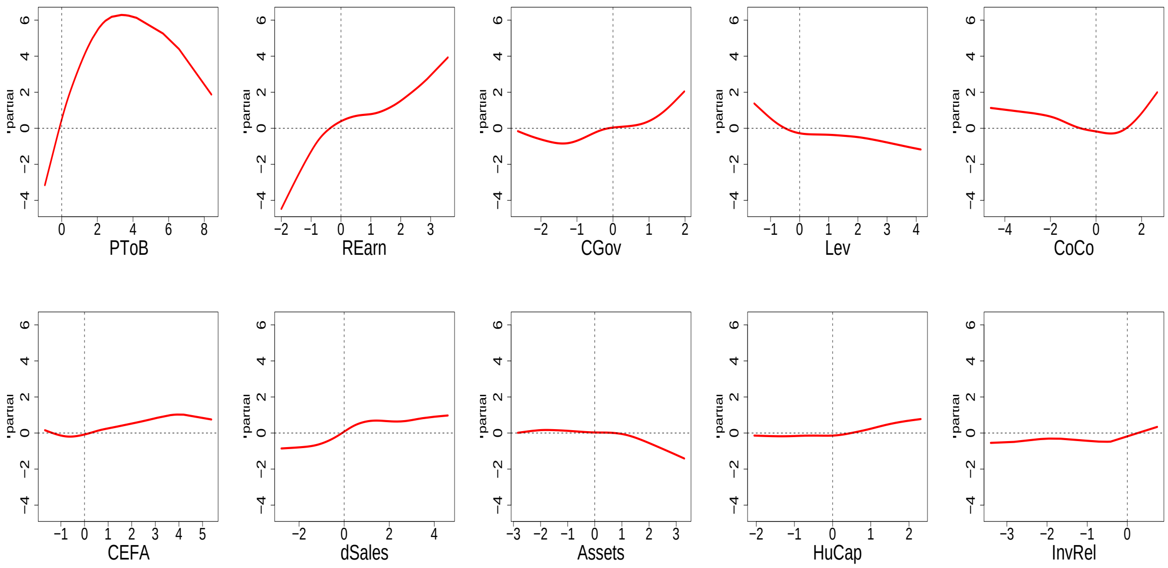

FIGURE 5. The partial contribution of the explanatory variables estimated with component-wise smoothing spline. The method yields a reduction of 17% of the prediction error with respect to a linear regression with all variables.

stopping early (which is usually needed to avoid over-fitting) often does variable selection, i.e. some of the predictors will never be picked during boosting iterations.

The performance of the two non-linear models is evaluated based on their average prediction error on the validation samples of the cross-validation exercise. The results are presented in the second line of Table 3. Boosting with component-wise smooth spline method yields a reduction of 17% of the prediction error with respect to a linear regression that uses all explanatory variables. Bagging regression trees reduces the prediction error with more than 21%.

Figures 5 displays the partial dependence functions (introduced in sub-section 6.4) for the variables found to significant by boosting with component-wise smooth splines. Recall that the partial dependence function f_i represents the effect of the independent variable X_i on f after accounting for the average effects of all the other variables. *Only ten variables* were found to have an impact on the ROA. Since the explanatory variables are centered and standardized, the abscissa negative (positive) values correspond to an under-the-average (over-the-average, respectively) performance. We note the strong non-linearities in the price to book, retained earnings, and codes of conduct variables.

Figure 6 displays the partial dependence functions for the dependent variables described in Section 4.3 in the case of bagging regression trees. The variables are chosen according to the ranking in the graph of relative importance in Figure 8.2 as well as by the magnitude of the partial dependence function. The contribution of the variables ranked lower than the tenth place

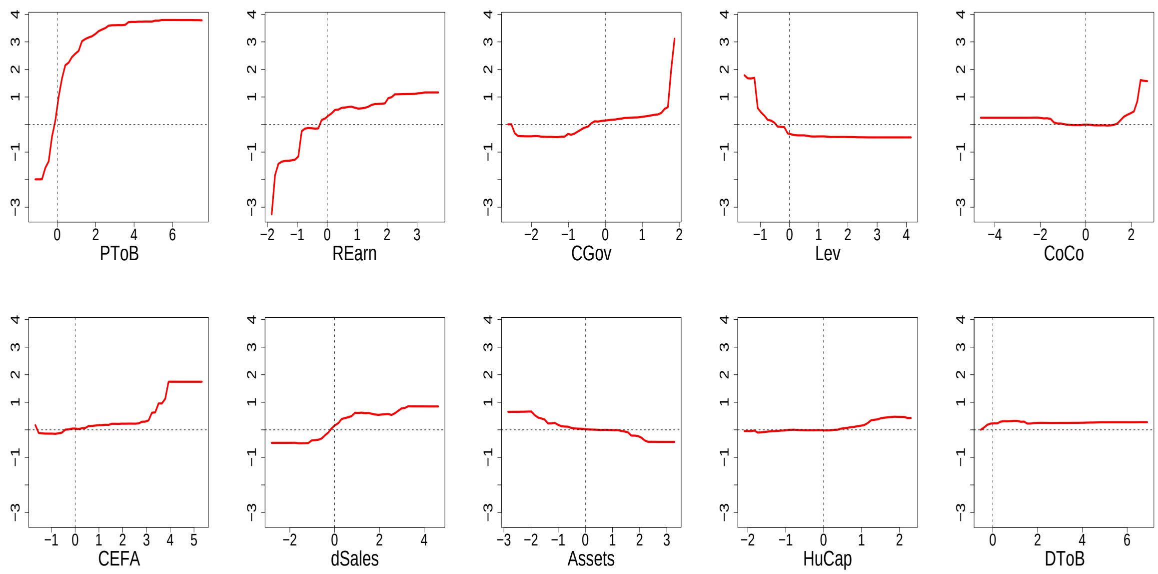

FIGURE 6. The partial contribution of the explanatory variables for ROA estimated by bagging trees. The method yields a reduction of 21% of the prediction error with respect to a linear regression that uses all explanatory variables.

is negligible. It is worth mentioning that the first nine most influential variables in explaining the profitability of the firm (as measured by ROA) identified by the method of bagging trees have also been chosen by the boosting smooth spline approach. It is worth noticing also that no industry dummy shows up among the significant variables.

A look at the graphs in Figures 5 and 6 show that the shape of the partial dependence functions estimated by the two extremely different non-linear methods, are very similar. However, some differences in the scale of the impact of various variables are to be noted. Strong non-linearities are evident in most of the variables. As the method of bagging regression trees seems to improve the prediction the most our interpretation of the results puts most weight on the graphs in Figure 6.

The estimated impact of the (significant) non-CSR variables on profitability is often confirmed by intuition. As expected a lower(higher)-than-average price to book value negatively (positively) influences the profitability, since the stock price is the present value of future estimated cash flows, which directly related to earnings. The main contribution of our analysis consists in estimating a strongly *non-linear* relationship. While small and moderate positive departures from the mean of the PtoB variable are associated with a strong increase in ROA (up to 2 to 4%), the effect tends to level off around 4% (and to even reverse according to the graph in Figure 5 where the maximum impact is of an extra 6% for firms that have a moderate above-average PToB) as the departures are more extreme. Retained earnings and leverage are increasingly and respectively decreasingly monotonely related to ROA as one would expect. They both are strongly monotone for less-than-average firms while leveling off on the positive side of the axis. Too little retained earnings can hurt ROA with up to 4% while small leverage can help it with up to 2%. Return on assets seems to depend monotoneously increasing on change in sales and capital intensity. The size of the firm seems negatively related to profitability as measured by ROA: smaller(bigger)-than-average firms seem a bit more (less) profitable (the size can change ROA with up to 0.5-1%). Turning now to the CSR variables we note that under-the-average Corporate Governance score seems to hurt profitability (subtracting up to 0.7%), although for firms that have extremely low scores the negative impact on profitability seems to go away. About average Corporate Governance scores do not help, while higher-than-average can raise the profitability with up to 2-3%. The impact of the Codes of Conduct scores is mixed with firms that display both well under-the-average and above-the-average scores benefiting a mere 0.25% on the negative side and a more substantial almost 2% on the positive side.

8.3. **The contribution of the Corporate Social Responsibility scores.** It is of course important to understand the impact the presence of the CSR variables might have on the five different procedures discussed as well as the magnitude of the contribution of the CSR variables to explaining return on assets. Towards this end the cross-validation exercise has been conducted using as explanatory variables only the non-CSR variables. The results are provided in Table 3, 4 and 5. All the tables display *averages* of specific statistics over the five repetitions in a five-fold cross-validation exercise.

Table 3 summarizes the behaviour of the five different methodologies on the two (related) data sets separately, i.e. the data set containing all the explanatory variables and the reduced one containing only the non-CSR variables.

Variables	Stepwise Regression	Breiman's Method	Boosting Linear	Boosting Spline	Bagging Trees
Without CSR	-1.3%	-0.6%	1.5%	11.9%	19.2%
All variables	1.3%	3.8%	4.8%	16.9%	21.2%

TABLE 3. Reduction in prediction error relative to a linear regression without any variable selection. First line reports the results obtained when the explanatory variables do not contained the CSR scores. The results on the second row correspond to using all available explanatory variables.

Independent of the set of dependent variables used, allowing for non-linearities in the relationship between profitability and explanatory variables as well as a carefully selection of (relevant) variables significantly improves predictability.

In the case when the CSR variables are excluded, choosing variables by step-wise algorithm or Breiman's approach increases actually the prediction error with 1.3% and 0.6%, respectively.

Remaining in the linear framework, a selection rule based on boosting with component-wise linear bases yield an improvement in prediction of 1.5%. The non-linear nature of the relationship between the dependent variables and the ROA is emphasized by the significant increase in the accuracy of predicting the measure of profitability with almost 12% in the case of boosting with component-wise smooth spline base and of 19.2% for bagged regression trees.

When the CSR variables are included both the step-wise and Breiman's rules improve the prediction error, the first with 1.3%, the second with 3.8%. In the case of the linear model a variable selection procedure based on boosting with component-wise linear bases is the best and improves the explanatory power by almost 5%. However, it is relaxing the assumption of linearity that brings, also in this case, the highest gains in terms of prediction. Boosting with component-wise smooth spline method yields a reduction of 17% of the prediction error with respect to a linear regression that uses all explanatory variables. Bagging regression trees reduces the prediction error with more than 21%. It is likely that the worse performance of the boosting with component-wise smooth spline is the price to pay for the more gratifying shapes of the curves in Figure 5. Bagging regression trees does not impose smoothness constrains on the shape of the function f. This extra flexibility is, possibly, responsible for the outstanding performance of the method.

The magnitude of the contribution of the CSR variables to explaining return on assets is given by the results in Table 4. Again, the analysis is model/method of variable selection specific. Overall, adding the CSR scores to the set of explanatory variables improves the prediction. However, the size of the improvement varies from -2% in the case of a linear model without any variable selection (in fact adding CSR variables without performing any variable selection worsens on average the prediction error) to 3.3% in the case of the spline (aided by boosting). We note that for the most performant approach, i.e. the regression trees, the gain is small (under 1%).

No selection Regression	Stepwise Regression	Breiman's Method	Boosting Linear	Boosting Spline	Bagging Trees
-2%	0.1%	2%	2.2%	3.3%	0.9%

TABLE 4. Reduction in prediction error due to inclusion of the CSR scores. The colons correspond to the different models/model selection approaches discussed in the paper: linear model with a) no, b) stepwise and c) Breiman's approach d) boosting variable selection followed by non-linear models e)splines and f) regression trees. The prediction error of the approach using only non-CSR variables is taken as reference (100%).

Finally, since most of the regression analysis are phrased in the language of the R^2 statistic, Table 5 reports this quantity for all the five approaches discussed (plus the linear regression without variable selection) both when all explanatory variables and when only non-CSR variables are used. We report the average R^2 for both the estimation sample, i.e. the four fifth of sample on which the model is selected and estimated, as well as for the validation sample, i.e. the other fifths of the entire sample that is used for model validation. As expected, for all the methods, the statistic is higher on the estimation than on the validation sample. An 'evaluation' based on the R^2 from the sample on which the model is estimated (the values on the rows with the label 'Estimation') would paint a totally wrong picture: first, the regression without variable selection would be the best linear model and second, it would be much closer to its non-linear competitors (boosting splines would be only 16% better in-sample as compared to 63% in out-of-sample comparison when no CSR variables are used).

Variables	Sample	No selection Regression	Stepwise Regression	Breiman's Method	Boosting Linear	Boosting Spline	Bagging Trees
Without CSR	Estimation	0.36	0.34	0.34	0.34	0.43	0.49
	Validation	0.25	0.23	0.25	0.27	0.33	0.38
All	Estimation	0.41	0.40	0.37	0.39	0.50	0.89
	Validation	0.22	0.23	0.25	0.29	0.35	0.39

TABLE 5. The average R^2 for both the estimation sample, i.e. the four fifth of sample on which the model is selected and estimated, as well as for the validation sample, i.e. the other fifths of the entire sample that is used for model validation. As expected, for all the methods the statistic is higher on the estimation than on the validation sample.

The overall picture emerged from the analysis based on prediction error remains unchanged when the measure of performance is the classical R^2. We see that the linear model, even when enhanced by theoretically sound variable selection methods, i.e. Breiman's approach, performs a lot worse than the non-linear models used. The regression tree approach seems to again slightly dominate the smooth spline method. It is worth mentioning that the linear regression without variable selection performs particularly poorly: adding the CSR variables worsens the performance.

9. Conclusions

Our analysis brings both good and bad news. We believe they are relevant for researchers and practitioners (corporate management, institutional investors and sustainability ranking companies) alike.

Let us start with the good news. First of all, we believe we demonstrated that non-linear methods bring important improvements in explaining profitability. Our findings are strong evidence that an increased level of statistical sophistication is needed if one wants to uncover subtle relationships explaining firm profitability. The relevance of our findings is buttressed by the fact that two very different non-linear approaches produce similar results delivering almost identical sets of relevant explanatory variable as well as significant reduction in the prediction error. We believe that an increased attention from the academic community to non-linear approaches might well be a source of important achievements. The time might have come when the linear regression battle horse is put to rest and other readily available more sophisticated (and more reasonable) statistical methods are given a chance to show what they can.

Second, besides known variables as price to book ratio, retained earnings, leverage, capital intensity, change in sales, size or dividend to book ratio, we find that a number of CSR variables like Corporate Governance and Codes of Conduct seem to have important explanatory power. In particular, the Corporate Governance positions ahead of common profitability explanatory variables as leverage, capital intensity, change in sales or size. Well above average scores in Corporate Governance can add as much as 3 to 4% to Return on Assets (depending on the method used). Only above-average 'price to book' variable or retained earnings can add more. Moreover, being a leader in Codes of Conduct can have a nearly double (positive) impact on ROA with respect to average change in sales (less than 2% increase in ROA relative to a 1% increase).

These findings are important for managers as they show that good Corporate Social Responsability performance can provide added value and should be taken into account by the corporate operational strategy. For sell-side analysts, they show that recommendations based on predicted earnings might benefit from taking into account some (chosen) CSR variables as well as from a more sophisticated statistical approach to prediction.

Let us move to the less good news (that leave however room for improvement in many ways). The first is that most of the CSR scores considered in the analysis do not seem to contribute to explaining profitability. The 'big absent' is the environment dimension. Contrary to many empirical results finding some (negative or positive) association between the CSR environmental dimension and firm performance (most evidence refers to financial performance measures rather than economic ones), our analysis shows no relation whatsoever. This finding might not be so surprising, recalling that the the three environmental scores we have used showed little deviation from maximum values (see Figure 8), a characteristic that "eco- efficiency" scores apparently share irrespective of the data provider.

The second less good news is that the contribution of the CSR variables accounts for little of the improvement in prediction error. It improves prediction by 1 to 3% (depending on the method used). The big gain in prediction reported in the paper is due to acknowledging the non-linear nature of the association between explanatory variables and the Return on Assets.

These negative results could be due to many reasons. It could be the sample we worked with. A larger and more diverse sample would have allowed for a more sensitive analysis and might have put into light contributions that, with the data at hand, are impossible to detect. The construction of the scores could provide, of course, another possible explanation. Aggregated differently, the answers to the questionnaires might have produces more informative scores. Or it could be that other statistical techniques can individuate better the set of relevant variables.

Most important, however, is to note that our findings do not suggest that the qualities these scores set to measure are not important. It's only to the extent to which the SAM scores measure correctly and completely the operational implementation of Corporate Social Responsability concepts and the Return on Assets truthfully reflects the profitability of a firm that the conclusions of our analysis carries to the the understanding of the 'true' association between Corporate Social Responsability and profitability. Most likely we have not yet found the best way to measure them and the results of the analysis might constitute a stimulus (for academics as well as for the rating companies) to rethinking the methodology behind the construction of such scores.

Acknowledgements: We thank Carl-Johan Francke and Christophe Churet of SAM for providing clean data and for a careful and prompt reading of the manuscript, to E. Carlsson for his instrumental help in getting the data and for many fruitful discussions on it and to T. Hothorn for promptly answering all our R questions.

10. Appendix: A more formal view on boosting

Boosting is a way of fitting an additive expansion in a set of elementary 'basis' functions. Generally it takes the form

$$f(x) = \sum_{m=1} \beta_m b(x; \gamma_m) \tag{10.6}$$

where β_m, $m = 1, \ldots, M$ are the expansion coefficients and $b(x; \gamma)$ are usually simple functions of the multivariate argument x, characterized by a set of parameters γ. The structural properties of the boosting function estimator are induced by a linear combination of the structural characteristics of the basis functions.

To make precise the estimator (10.6) one needs to specify the basis functions and a modality of estimation of the coefficients β.

Forward stage-wise modeling approximates the solution of (10.6) by sequentially adding new basis functions to the expansion without adjusting the parameters or the coefficients of those that have already been added. More concretely, the steps are

(1) Initialize $f_0(x) = 0$.
(2) For $m = 1$ to M
   (a) Compute
   $$(\beta_m, \gamma_m) = \operatorname*{argmin}_{\beta,\gamma} \sum_{i=1} L(y_i, f_{m-1}(x_i) + \beta b(x_i; \gamma)).$$
   (b) Set
   $$f_m(x) := f_{m-1}(x_i) + \beta_m b(x_i; \gamma_m).$$

The loss function L will be for us the squared-error loss

$$L(y, f(x)) = (y - f(x))^2. \tag{10.7}$$

Under this assumption the previous algorithm turns into

(1) Initialize $f_0(x) = 0$.

(2) For $m = 1$ to M

(a) Compute the residuals

$$U_i = Y_i - f_{m-1}(X_i), \quad i = 1, \ldots, n.$$

Fit the residual vector $U_1, \ldots, U_n$ to $X_1, \ldots, X_n$ (by some type of regression) to estimate (β_m, γ_m).

(b) Set

$$f_m(x) := f_{m-1}(x_i) + \beta_m b(x_i; \gamma_m).$$

In other words, the L_2-boosting amounts to refitting residuals multiple times.

Estimation of f with boosting is usually done by pursuing an iterative steepest descent of the empirical risk $n^{-1} \sum_{i=1^n L(Y_i, f(X_i))}$. To fully specify our approach we need to still define the basis functions to be used in the sequel. They correspond to component wise linear least squares and component wise smoothing spline. More formally they are

$$\hat{\beta}^{(\mathcal{S})} x^{(\mathcal{S})} \tag{10.8}$$

$$\hat{\beta}^{(j)} = \sum_{i=1}^{n} X_i^{(j)} U_i / \sum_{i=1}^{n} (X_i^{(j)})^2, \quad \mathcal{S} = \text{argmin} \sum_{i=1}^{n} (U_i - \hat{\beta}^{(j)} X_i^{(j)})^2$$

(10.9)

and

$$h^{(\mathcal{S})}(x^{(\mathcal{S})}) \tag{10.10}$$

$$h^{(j)} = \underset{h}{\text{argmin}} \sum_{i=1}^{n} (U_i - h(X_i^{(j)}))^2 + \lambda \int (h^{''})^2 dx, \quad \mathcal{S} = \text{argmin} \sum_{i=1}^{n} (U_i - h^{(j)}(X_i^{(j)}))^2,$$

where λ is a tuning parameter.

The basis (10.8) selects in every iteration the best one predictor variable in a simple linear model (not necessary a different one in each iteration) in the sense of ordinary least square fitting. The method is also known as matching pursuit. The component wise smoothing spline (10.10) belongs to the class of generalized linear models. The best one predictor variable is chosen and its partial contribution to f is estimated by a least squares smoothing spline. The degrees of freedom in the smoothing spline base procedure are chosen small such as df=4. This choice yields low variance but possibly large bias which can be typically removed by additional boosting iterations.

References

[1] Allen, D. M., 1974. The relationship between variable selection and data augmentation and a method for prediction. Technometrics, 16, 125-127.

[2] Barnett M.,and Salomon R., 2006. Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance. Strategic Management journal, 27, 1101-1122.

[3] Bauer, R., Koedijk K., and Otten R., 2005. International evidence on ethical mutual fund performance and investment style. Journal of Banking and Finance, 29, 1751-1767.
[4] Brammer S., Brooks C., and Pavelin S., 2006. Corporate social performance and stock returns: UK evidence from disaggregated measures. Financial Management, 35(3), 97-116.
[5] Breiman, L., 1992. The little bootstrap and other related methods for dimensionality selection in regression: X-fixed prediction error. Journal of the American Statistical Association, 87, 738-754.
[6] Breiman, L., 1996a. Bagging predictors. Machine Learning 24, 123-140.
[7] Breiman, L., 1996b. Heuristics of instability and stabilization in model selection. Ann. Statist. 24, 2350-2383.
[8] Breiman, L., and Spector, P., 1992. Sub-model selection and evaluation in regression; The X-random case. International Statistical review, 60(3), 291-319.
[9] Bühlmann, P., and Yu, B., 2003. Boosting With the L2 Loss: Regression and Classification. Journal of the American Statistical Association, 98, 324-339.
[10] Burke, L., Logsdon J.M., Mitchell W., Reiner M., and Vogel D., 1986. Corporate community involvement in the San Francisco bay area. California Management Review, XXVIII(3), 122-141.
[11] Capon, N., Farley, J.U., and Hoenig, S., 1990. Determinants of financial performance: A meta-analysis. Management Science, 36, 1143-1159.
[12] Chatfield, C., 1995. Model uncertainty, data mining and statistical inference. Journal of the Royal Statistical Society, 158, 419-466.
[13] Cohen, M.A., Fenn, S.A. , and Naimon, J., 1995. Environmental and Financial Performance: Are they related? Investor Responsibility Research Center.
[14] Cohen, M.A., Fenn, S.A., and Konar., S., 1997. Environmental and Financial Performance: Are They Related? Working paper.
[15] Derwall, J., Güenster N., Bauer, R., and Koedijk, K., 2005. The eco-efficiency premium puzzle. Financial Analysts Journal, 61(2), 51-63.
[16] DiBartolomeo, D., and Kurtz, L., 1999. Managing risk exposures of socially screened portfolios. Northfield Information Services.
[17] Dillenburg, S., Greene, T., and Erekson, H.O., 2003. Approaching social responsible investment with a comprehensive ratings scheme: total social impact. Journal of Business Ethics, 43, 167-177.
[18] Diltz, J.D., 1995. Does social screening affect portfolio performance? The Journal of Investing, 4(1), 64-69.
[19] Dowell, G., Hart,S., and Yeung,B., 2000. Do corporate global environmental standards create or destroy value? Management Science, 46(8), 1059-1074.
[20] Efron, B., 1986. How biased is the apparent error rate of a prediction rule? Journal of the American Statistical Association, 81, 461-470.
[21] Fama, E.F., and French, K.R., 1999. Disappearing dividends: Changing firm characteristics or increased reluctance to pay? Unpublished manuscript. University of Chicago, Graduate School of Business.
[22] Fama, E.F.,and French, K.R., 2000. Forecasting profitability and earnings. Journal of Business, 73(2), 161-175.
[23] Feldman, S.J., Soyka, P.A., and Ameer, P.G., 1997. Does improving a firm's environmental management system and environmental performance result in a higher stock price? Journal of Investing, 6(4), 87-97.
[24] Freedman, D. A., and Peters, S. C., 1984. Bootstrapping a regression equation: Some empirical results. Journal of the American Statistical Association, 79, 97-106.
[25] Freund, Y., and Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. System Sci. 55, 119-139.
[26] Geisser, S., 1975. The predictive sample reuse method with applications. Journal of the American Statistical Association, 70,320-328.
[27] Güenster, N., Derwall, J., Bauer, R., and Koedijk, K., 2005. The economic value of corporate eco-efficiency. Academy of Management Conference. Available at SSRN: http://ssrn.com/abstract=675628.
[28] Guerard, J.B., Jr., 1997. Additional evidence on the cost of being socially responsible in investing. Journal of Investing, 6(4), 31-35.

[29] Hamilton, J. T., 1995. Pollution as news: Media and stock market reactions to the toxic release inventory data. Journal of Environmental Economics and Management, 28, 98-113.
[30] Hamilton, S.,Jo, H., Statman.,M., 1993. Doing Well While Doing Good? The Investment Performance of Socially Responsible Mutual Funds. Financial Analysts Journal, 49(6), 62-66.
[31] Hart, S., and Ahuja.,G., 1996. Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. Business Strategy and the Environment, 5, 30-37.
[32] Hastie, T and Tibshirani, T., 1986. Generalized Additive Models. Statistical Science, 1, 297-318.
[33] Heinze, D.C., 1976. Financial correlates of a social involvement measure. Akron Business and Economic Review, 7(1), 48-51.
[34] Herremans, I., Akathaporn, P., and McInnes, M., 1993. An Investigation of Corporate Social Responsibility Reputation and Economic Performance Accounting, Organizations and Society, 18, 587-604.
[35] Hocking, R.R., 1976. The analysis and selection of variables in linear regression. Biometrics, 32, 1-49.
[36] Karpoff, J., Lott, J., and Rankine,G., 1998. Environmental violations, legal penalties, and reputation costs. John M. Ohlin Law and Economics Working paper, 71.
[37] Kinder, P. T., and A. Domini, 1998. Social Screening: Paradigms Old and New. The Investment Research Guide to Socially Responsible Investing (Investment Research Forums, Plano, TX), pp. 10-18.
[38] King A., and Lenox M., 2001. Does it really pay to be green? An empirical study of firm environmental and financial performance. Journal of Industrial Ecology, 5(1), 105-116.
[39] Luther, R.G., Matatko, J., 1994. The performance of ethical unit trusts: Choosing an appropriate benchmark. British Accounting Review, 26, 77-89.
[40] McGuire, J. B., Sundgren, A., and Schneeweis,T., 1988. Corporate social responsibility and firm financial performance. Academy of Management Journal, 31(4), 854-872.
[41] McWilliams, A., and D. Siegel, 2000. Corporate social responsibility and financial performance: Correlation or misspecification? Strategic Management Journal, 21(5), 603-609.
[42] Miller, A. J., 1984. Selection of subsets of regression variables. Journal of the Royal Statistical Society, 147, 389-425.
[43] Miller, A. J., 1990. Subset Selection in Regression. Chapman and Hall, New York.
[44] Miller, M. and Modigliani, F., 1961. Dividend Policy, Growth, and the Valuation of Shares. The Journal of Business, 34, 411-433.
[45] Modigliani, F., Miller, M. H., 1958. The cost of capital, corporation finance and the theory of investment. American Economic Review 48, 261-297.
[46] Muoghalu, M., Robinson, H.D., and Glascock.,J., 1990. Hazardous waste lawsuits, stockholder returns, and deterrence. Southern Economic Journal, 57, 357-370.
[47] Moskowitz, M., 1972. Choosing socially responsible stocks. Business and Society Review, 1, 71-75.
[48] Mosteller, F. and J. W. Tukey, 1977. Data Analysis and Regression. Addison-Wesley, Reading, Massachusetts.
[49] Palmer K.P., Oates W.E., and Portney P.R., 1995. Tightening environmental standards: the benefit-cost or no cost paradigm. Journal of Economic Perspectives, 9(4), 119-132.
[50] Porter M., and van der Linde C., 1995. Green and competitive: ending the stalemate. Harvard Business Review, Sept-Oct, 120-134.
[51] Preston, L. E.,and OBannon, D. P., 1997. The corporate social-financial performance relationship: A typology and analysis. Business and Society 36, 419-429.
[52] Russo, M.V., and Fouts, P.A., 1997. A resource-based perspective on corporate environmental performance and profitability. Academy of Management Journal, 40(3), 534-559.
[53] Schaltegger, S., and Müller, K., 1998. Calculating the true profitability of pollution prevention. In Benett M., and James P., (Eds.), The Green Bottom Line: Environmental Accounting for Management; Current Practice and Future Trends. Greenleaf Publishing, 86-99.
[54] Shane, P.B., and Spicer, B.H., 1983. Market Response to Environmental Information Produced Outside the Firm. The Accounting Review, 58(3), 521-285.
[55] Shao, J., 1996. Bootstrap model selection. Journal of the American Statistical Association, 91, 655-665.

[56] Siebert, H., Eichberger J., Gronych R., and Pethig R., 1980. Trade and environment: a theoretical enquiry. Amsterdam: Elsevier Scientific Publishing Company.

[57] Social Investment Forum, 2005. Report on socially responsible investing trends in the United States. Available as www.sbnow.org/doc/SRITrendsReport2005PPPresentation.pdf

[58] Statman, M., 2000. Socially responsible mutual funds. Financial Analysts Journal, 56(3), 30-39.

[59] Steinberg, D., and Colla, P., 1995. CART: Tree-Structured Nonparametric Data Analysis. Salford Systems, San Diego, CA.

[60] Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, 36, 111-133.

[61] Ullmann, A., 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. The Academy of Management Review, 10 (3), 540-557.

[62] United Nations Environment Programme Finance Initiative, United Kingdom Social Investment Forum and Sustainable Pensions Project Joint Report, 2007.

[63] Vance, S. C., 1975. Are socially responsible corporations good investment risks? Management Review, 64, 18-24.

[64] Waddock, S. A., and Graves S., 1997. The corporate social performance-financial performance link. Strategic Management Journal, 18 (4), 303-319.

[65] Wagner, M., Phu, N.V., Azomahou, T., and Wehrmeyer, W., 2002. The relationship between the environmental and economic performance of firms: and empirical analysis of the European paper industry. Corporate Social Responsibility and Environmental Management, 9, 133-146.

[66] World Commission on Environment and Development, 1987. Our common future. Oxford University Press