



**HOW TO MEASURE CORPORATE SOCIAL
RESPONSIBILITY**

**Marco NICOLosi — Stefano GRASSI
Elena STANGHELLINI**

Quaderno n. 96 — Ottobre 2011

**QUADERNI DEL DIPARTIMENTO
DI ECONOMIA, FINANZA
E STATISTICA**

How to measure Corporate Social Responsibility*

Marco Nicolosi[†], Stefano Grassi[‡], Elena Stanghellini[§]

October 31, 2011

Abstract

Compliance with Corporate Social Responsibility (CSR) standards may require capacity that varies from one aspect to the other and companies in different industries may encounter different difficulties. Since CSR is a multidimensional concept, latent variable models may be usefully employed to provide a unidimensional measure of the ability of a firm to fulfil CSR standards. A methodology based on Item Response Theory has been implemented on the KLD sustainability dataset. Results show that companies in the industries Oil & Gas, Industrials, Basic Materials and Telecommunications have a higher difficulty to meet the CSR standards. Criteria based on Environment, Community relations and Product quality have a large capacity to select the firms with the best CSR performance, while Governance does not exhibit similar behavior. A stock selection based on the ranking of the firms according to our CSR measure outperforms, in terms of risk-adjusted returns, stock selection based on other criteria.

Keywords: Socially Responsible Investment; CSR ability; latent variable model; item response theory

Jel Classification: C43; G11; G14; G19

*This study has been funded by the Swedish Foundation For Strategic Environmental Research, MISTRA.

[†]University of Perugia, Department EFS, Via A. Pascoli 20 - 06123 Perugia, Italy; e-mail address: nicolosi@stat.unipg.it

[‡]Aarhus University, CREATES, Bartholins Alle 10, 1322 University Park, DK-8000 Aarhus C, Denmark; e-mail address: sgrassi@creates.au.dk

[§]University of Perugia, Department EFS, Via A. Pascoli 20 - 06123 Perugia, Italy; e-mail address: elena.stanghellini@stat.unipg.it

1 Introduction

Socially Responsible Investment (SRI) has become a widespread practise in most industrialized countries. Estimates of the Social Investment Forum indicate that in 2010 about 12.2% of the total assets in US was invested along this line. Academic research has focused mainly on establishing a link between Corporate Social Responsibility (CSR) and financial performance, both at corporate and at portfolio level. Theoretical arguments exist in favor of a negative associations as well as a positive one or a neutral one.

Empirical research has found conflicting evidence. After controlling for investment style, Bauer et al. 2005 found no significant difference in risk-adjusted returns between ethical and conventional funds. However, the ethical funds tend to be less exposed to market variability of returns and more growth-oriented than conventional ones. The performance of ethical funds has increased over time, and this may be due to an initial learning phase. Similar evolution has been found by Barnett and Salomon (2006), who note that there is a variation between SRI in terms of types and intensity of screening. Different types of social screening may have different impact on returns. As an instance, funds that excluded firms that violated norms of equally employment suffered a financial penalty and so did the funds that selected firms that comply with environmental standards. On the other hand, funds that include firms that fostered positive relationship with their social community outperformed the others.

The impact of different socially responsible screening policies on the mean-variance optimal portfolios has been considered in Herzel et al. (2011). Their findings, based on KLD Research and Analytics rating system over the period 1993-2008, suggest that CSR screening has a small impact in terms of Sharpe ratio even though it has a great impact in terms of market capitalization. They notice further that if short selling is not allowed, the diversification opportunities decrease significantly only in the case of a screening based on environmental criteria.

Derwall et al. (2011) provided a segmentation of the socially responsible investors into values-driven and profit-oriented ones. While investors of the first group are able to sacrifice financial returns to derive non-pecuniary utility, and are therefore served by negative screens that avoid controversial stocks, investors belonging to the second group are motivated by the assumption that stocks with the best performing firms in CSR standards also produce superior returns. According to the shunned-stock-hypothesis, the

exclusion of controversial stocks by the values-driven investors pushes down their prices producing superior abnormal returns. On the other hand, the errors-in-expectations hypothesis explains the superior returns of the socially responsible stocks in terms of a low reaction of the market to recognize the positive impact of CSR practice on future cash flows. Since SRI mutual funds adopt a mixture of screens, the segmentation may explain why they neither outperform nor underperform conventional peers.

Socially responsible investors are mainly interested in implementing investment strategies that embody criteria other than financial risk and return. After performing some negative screening, the usual way to take the CSR information into account in the decision process on investments is to rank firms according to some index measuring their CSR performance and select portfolios accordingly. Several different criteria have appeared in the literature. Kempf and Osthoff (2007) formed portfolios with companies that score high and low on some CSR characteristics. Based on KLD rates, they found that a portfolio composed by the 10 percent of companies with the strongest Employee relations or Community involvement outperformed the one composed by the worst 10 percent. The highest abnormal returns were obtained by taking the portfolio of the stocks with the highest average rating of all the considered CSR characteristics and employing a best-in-class approach in order to account for the industry effects.

Similarly, Statman and Glushkov (2009) constructed portfolios by taking the best and worst companies ranked by an industry-adjusted score. For each characteristic accounted by KLD, the industry-adjusted score of a firm is computed as the difference between the firm's score in that characteristic and the average score of all companies in the same industry. They found that the portfolios of high ranked stocks in the Community, Employee relations or Environment yielded higher returns than portfolios of low ranked stocks.

Although a quantitative univariate syntheses of the available information is a desirable requirement, there is a general agreement that the average is not the best way to summarize the data, see e.g. Hopkins (2005) who after comparing six different known measurement frameworks concluded that "the power of the average seems to hide a variety of sins". In line with this, Derwall et al. (2011) comment that "CSR is a multidimensional and partially subjective concept, and investors lack the tools needed to adequately measure CSR practices and their effect on the fundamental value of the firm".

This paper addresses the issue on how to build an aggregate measure of CSR. We propose to use Item Response Models (see De Boeck and Wilson,

2003) to extract a unidimensional score, which we call CSR ability, that captures the capacity of a firm to comply with the different CSR standards. Item response analysis originates in psychometric literature, where the aim is to measure the ability of an individual to answer correctly to a series of questions. In this framework, the ability is a latent trait that is measured by a set of items of a questionnaire: these are binary indicators taking value 1 if the answer is correct and 0 otherwise. The model allows to take into account that different items have a different degree of difficulty and therefore a value 1 on a difficult item should be given more weight than a value 1 on an easy one. Furthermore, different items can have a different power to discriminate among subjects. For example, an item to which all the individuals, independently from their level of ability, give the same answer is a non discriminant item. Moreover, the difficulty of an item can vary with the age, the gender or the social class of an individual. It then follows that (a) different patterns of responses are not exchangeable and (b) the same pattern of responses has a different value depending on the personal features of the respondent. All these aspects can be accounted for within this class of models. The ability of a subject is measured through an item response model that weights the responses differently according to the different difficulties of the items as well as the specific group to which the subject belongs. Later extensions, that we used here, include the possibility to deal with data that are ordered responses.

We think of the companies as the subjects while the different CSR aspects as the items. The item response model has been implemented yearly from 1992 to 2008 on the KLD rating system, that measures the CSR performances of firms in the US with respect to seven dimensions capturing Corporate Governance, Environment and Social issues. The firms' ability to comply with CSR standards are extracted via this model that weights each CSR aspect differently, with weights that are allowed to vary across industries. As a matter of fact, compliance with CSR standards may imply additional costs that vary across industries or sectors and any analysis that does not explicitly take this issue into account is confounded (see Benson et al. 2006). Results show that companies in Oil & Gas as well as in Industrials, Basic Materials and Telecommunications have a higher difficulty to comply with the CSR standards. For what concerns the items, criteria based on Environment, Community and Product quality have a large capacity to select the firms with the best CSR performance and therefore the ability is more influenced by these aspects, while Governance is not a discriminant item.

As an application, we used the CSR ability to rank companies and perform portfolio selection. We constructed high and low ranked portfolios (equally weighted, value weighted and mean-variance optimal) according to the proposed measure. We also constructed the high and low ranked portfolios according to a benchmark ordering based on the raw average of the CSR scores over the different CSR dimensions. Finally we computed the risk-adjusted returns as measured by the Jensen's α in different factor models, including the Carhart's model (Carhart, 1997) and an industry model (similarly to Huij and Derwall, 2011). We found that the high ranked portfolios according to the proposed ordering outperformed the low ranked ones. Moreover the high ranked portfolios according to the CSR ability outperformed the high ranked portfolios according to the row average of the CSR scores.

Notice that our proposal is mainly methodological, and aims to foster the use of theoretically robust methods, based on latent variable models, to provide a unidimensional measure that captures CSR ability. Approaches based on latent variables can be implemented on other rating databases with different measurement schemes.

The rest of the paper is organized as follows: in Section 2 an introduction of the latent variable models is given while in Section 3 a description of the CSR data and the construction of the ordinal responses is detailed. The polytomous item response model we implemented is described in Section 4 while in Section 5 the qualitative findings given by the model are reviewed. In Section 6 an application to portfolio selection is illustrated and in Section 7 some conclusions are presented.

2 Latent variable models

In many educational and psychological measurement situations there is an underlying variable of interest. This variable is often something that is intuitively understood, such as intelligence or attitude, but it cannot be measured directly as can height or weight, for example, since it is a concept rather than a physical dimension. This is what psychometricians refer to as an unobservable, or latent, trait. Such a variable is easily described via its attributes, which altogether constitute partial and imperfect measurements. A review on latent variable models can be found in Skrondal and Rabe-Heskett (2004).

A primary goal of educational and psychological measurement is the determination of how much of such a latent trait a person possesses. Since

most of the research has dealt with variables such as scholastic, mathematical or language skills, the generic term "ability" is used in this context to refer to such latent trait. For this purpose, it is necessary to have a scale of measurement, a ruler having a given metric. This can be for example a set of questions, or items, with binary or multiple answers.

Items may possess different capacity to discriminate among people. For example, all subjects tend to give correct answers to a trivial item or wrong answers to a difficult one. In both cases, this item is not discriminant. Furthermore, different background variables, such as gender or social class, have to be taken into consideration, as their effect may be in the direction of making an item more complex to some people than for others. If so, a correct answer provided by a subject of one group may require a higher level of ability than a correct answer provided by a subject of another group.

A correct way to extract the ability of subjects has to give different weights to each item and to take the effect of background variables into account. In our context, we want to extract the ability of a firm to fulfil the sustainability standards. Corporate Social Responsibility has a complex structure, usually measured by several dimensions. If we substitute a generic person with a generic firm, and we consider each dimension as an item of a measurement model, after introducing covariates to take the effects of industries into account, latent variable models can be successfully applied to measure the social responsibility of a firm.

3 Sustainability scores

In this section we define our metric (set of questions) for the latent variable models. We use the dataset released by KLD Research and Analytics¹. KLD rates US companies for what concerns the following seven dimensions: Governance, Community, Diversity, Employee relations, Environment, Human rights, Product quality. For each of these dimensions KLD considers different qualitative binary indicators taking values 0 or 1. There are two types of indicators: "strength" and "concern". A score 1 in a strength is "positive", meaning that the company has a proactive behavior in complying with the standards; on the other hand, a score 1 in a concern indicator has to be con-

¹KLD Research and Analytics was acquired by RiskMetrics at the end of 2009. The expertise developed by KLD over the past 20 years was integrated in the MSCI ESG Research, through the MSCI's recent acquisition of RiskMetrics.

sidered as “negative”, indicating a weakness of the company to satisfy the standards. A company with a score equal to 0 in any strength and concern of a given dimension cannot be qualified for that aspect. In addition, KLD also provides negative ratings on controversial business area such as Alcohol, Gambling, Firearms, Military, Nuclear Power and Tobacco.

The KLD’s sustainability scores were assigned on the basis of the company’s corporate social responsibility reports and public information, and after a direct engagement. KLD released ratings yearly. The scores reflecting the performances of a certain year were published at the beginning of the following year. From 1991 to 2000, KLD covered approximately 650 companies belonging to the Domini 400 Social index² and/or to the S&P500 index. In 2001 KLD expanded its coverage to include the largest 1000 US companies by market capitalization. Since 2003 KLD has provided ratings for the largest 3000 US firms. KLD has used the names and, since 1995, the CUSIP codes in order to identify the companies.

In our analysis we considered the years from 1992 to 2008 and an investment universe consisting of a comprehensive subset of the companies belonging to the S&P500 index and/or to the MSCI KLD 400 Social index³. Since each item involves a set of binary indicators, we aggregated strengths and concerns, in order to reduce the dimensionality of the problem. As in Manescu (2011), for each company p in our universe of investment and for each item $i = 1, \dots, 7$, yearly we built the net score S_{pi} , as the difference between the average value over the strengths and the average value over the concerns:

$$S_{pi} = \frac{\sum_{k=1}^{n_i} s_{pi}^k}{n_i} - \frac{\sum_{k=1}^{m_i} c_{pi}^k}{m_i} \quad (3.1)$$

where s_{pi}^k (c_{pi}^k) is the value of the binary k -th indicator measuring the strength (concern) of corporate p in the item-dimension i , and m_i (n_i) is the number of binary indicators measuring the strengths (concerns) of that item-dimension. The m_i and n_i may vary for each item and with time. Therefore, averaging over the number of strengths and concerns in the computation of the net score S_{pi} makes all the items comparable. Finally we built the categorically ordered response variable Y_{pi} as follows:

²The Domini 400 Social index is now called MSCI KLD 400 Social index.

³Financial data were downloaded from Datastream that uses ISIN codes to identify the companies. Matching the financial data with the KLD data, identified through the names or the CUSIP codes, produced a loss of about 5% in the total market capitalization of our investment universe as some companies of the KLD dataset were not identifiable.

- (i) $Y_{pi} = -1$ if $S_{pi} < 0$
- (ii) $Y_{pi} = 0$ if $S_{pi} = 0$
- (iii) $Y_{pi} = 1$ if $S_{pi} > 0$

We therefore considered only three categories $u = -1, 0, 1$ for the ordered responses Y_{pi} . As an example, $Y_{pi} = -1$ means that the company p raised more concerns than strengths on item i . This is not the unique choice we could adopt, but it turned out to be the best solution. Considering more categories leads to a very unbalanced distribution of the responses which gives rise to a lot of computational difficulties in the estimation, and sometimes even to a non-identifiability of the parameters of the latent model.

4 Polytomous item response models

We here introduce the class of latent variable models used for our purposes, i.e. the polytomous item response models (see De Boeck and Wilson, 2003). The starting point is the set of categorically ordered responses (or scores) Y_{pi} built as in the previous section. For each firm p , we modeled these answers as expressions of a latent dimension $\tilde{\eta}_p$, measuring the firm's social responsibility. The variable $\tilde{\eta}_p$ is supposed to be a stochastic variable whose prior distribution is normal with zero mean and an unknown variance.

Item response models give an expression of the probability of an individual p to have a score Y_{pi} in the item i not greater than category u , $u \in \{-1, 0, 1\}$ conditionally to his/her latent ability $\tilde{\eta}_p$. Mathematically this is

$$P(Y_{pi} \leq u \mid \tilde{\eta}_p) = G(f(\tilde{\eta}_p))$$

where $g = G^{-1}$ is referred to as the link function, connecting the categorical inputs Y_{pi} to the latent variable $\tilde{\eta}_p$. Function f is a linear function of $\tilde{\eta}_p$ that has to be specified in order to set the model. This function accounts for the capability of discrimination of an item and for the difficulty that companies in different industries may have to comply with different aspects of CSR. According to the Industrial Classification Benchmark (ICB), the following industries have been considered: Basic Materials, Consumer Goods, Consumer Services, Financials, Healthcare, Industrials, Oil & Gas, Technology, Telecommunications, Utilities.

In our work we used the Probit⁴ for the link function g . The following specification was allowed for f :

$$f(\tilde{\eta}_p) = k_u + b_{iu} - \left(\sum_r \beta_r D_p^r + \tilde{\eta}_p \right).$$

Coefficients k_u are thresholds that differentiate between categories. Such thresholds are modified by the b_{iu} parameters, depending both on item i and category u . Notice that $k_1 = \infty$ and $b_{i1} = 0$ for all the items by construction. Given a category u , $k_u + b_{iu}$ represents the difficulty to obtain a score higher than u in item i . Difficult items have large b_{iu} . The binary indicator D_p^r takes value 1 if firm p belongs to industry r , where label r runs over the industries. Therefore, the difficulty is modified by the industry effect represented by the coefficient β_r . Small values of β_r indicate that firms in industry r have higher difficulty in getting high scores than the others. The capability of discrimination of an item i depends on the difference $(k_{u'} + b_{iu'}) - (k_u + b_{iu})$, with $u' = 0$ and $u = -1$, which we call discriminatory coefficient. Higher values of this difference give higher probabilities that an individual p with a large (small) value of the latent ability has a high (low) score Y_{pi} .

The model is estimated by maximization of the marginal likelihood. Identification constraints impose that, for one industry r , $\beta_r = 0$ and, for one item i , $b_{iu} = 0$ for all categories u .

To give a hint of the implication of the proposed model, in Figure 1 the estimated probability of a score u with respect to the ability for item (a) Employment and (b) Community for companies in the industry Financials in year 2007 is presented. The first one is a non discriminating item, as for a wide range of ability around zero there is a rather substantial probability of each category u ; on the contrary the second one is a discriminating item, as low values of the ability give a high probability of $u = -1$, values around zero of the ability give a high probability of $u = 0$ and high values of the ability give a high probability of $u = 1$.

In Figure 2 the probability of a score 1 in item Environment for different industries for year 2007 is reported. It shows that, in that year firms in Oil & Gas have the greatest difficulty to present such a score, while firms in industry Technology have the lowest difficulty. In Figure 3 the estimated probability in year 2007 of a score 1 in all items for industry Financials is presented.

⁴It is the inverse of the standard normal cumulative distribution function. Another popular choice for the link function is the Logit function.

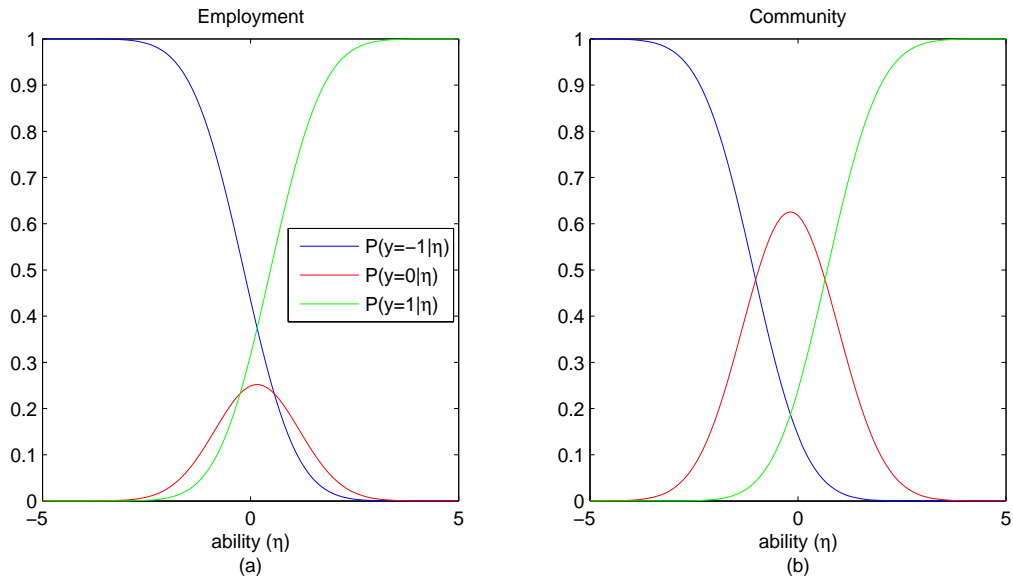


Figure 1: Conditional probabilities. Estimated probability of a score $u \in \{-1, 0, 1\}$ as a function of the ability for item (a) Employment and (b) Community for companies in industry Financials (year 2007)

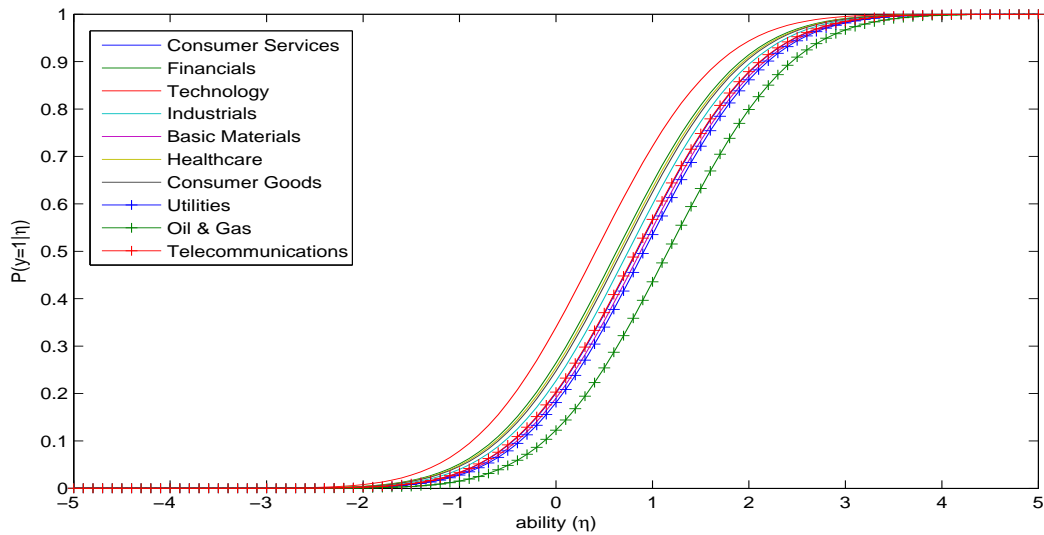


Figure 2: Comparison of industries. Probability of a score 1 in item Environment for different industries (year 2007)

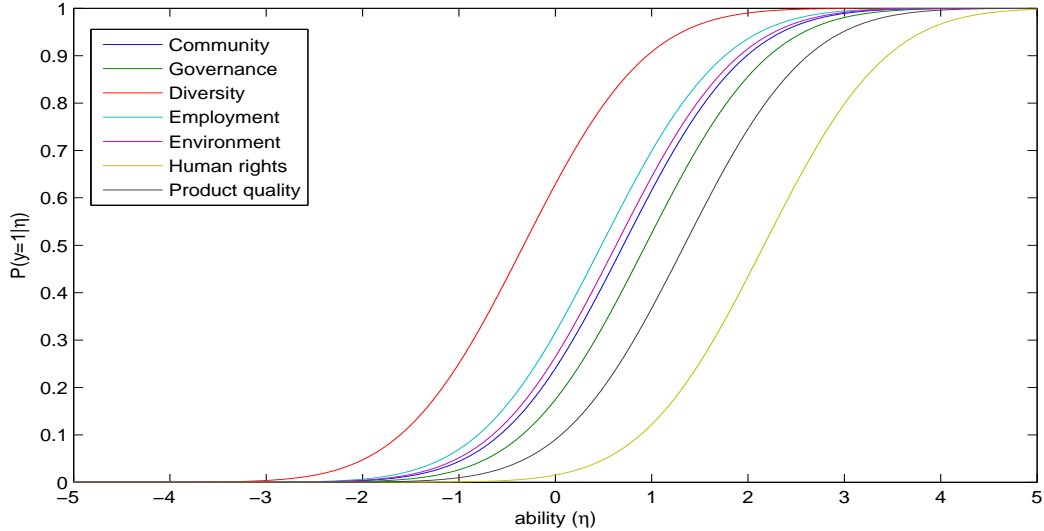


Figure 3: Comparison of items. Estimated probability of a score 1 in all items for industry Financials (year 2007)

The ordering of the items in terms of their difficulty for $u = 1$ shows that Diversity is the easiest item while Human rights is the most difficult one.

Once the model is estimated, following the Bayes' formula, the posterior distribution of the latent variable $\tilde{\eta}_p$ is computed. The expected value of this posterior distribution is taken as the CSR ability for company p , that we indicate with η_p . This is a one dimensional variable that provides a synthesis of the multidimensional nature of CSR behavior. It is built in a way that takes the industry effects and the different discriminant power of each item into account. The largest contribution to the ability is given from the most discriminant items.

5 Discussion of the IRM results

In this section we look at the evolution over time of the relevant parameters of the model. Figure 4 shows the evolution of β_r . Industry Oil & Gas consistently presents the highest difficulty to meet CSR standards. Industrials, Basic Materials and Telecommunications show a higher difficulty than the reference industry, Consumer Services, contrary to what exhibited by the

industry Financials.

Figure 5 (upper panel) shows the evolution of the difficulty of a score 1 for each item i while Figure 5 (lower panel) shows the evolution of the discriminatory coefficients. Both panels exhibit a rather stable behavior, especially after 1997.

Environment, Community and Product quality are rather discriminant items. This implies that the ability extracted from our model is strongly influenced by these dimensions. The impact of these CSR dimensions on financial performances has been extensively studied in literature. Derwall et al. (2005) showed that over the period 1995-2003 high ranked portfolios according to the Innovest Strategic Advisors' eco-efficiency scores provided higher average returns than the low ranked ones. Brammer et al. (2006) found that stock returns are negatively correlated to environmental (and employment) indicators while a small positive relation was found with a community indicator. In Manescu (2011) the author analyzed the period from 1992 to 2008, finding that only the Community dimension of the KLD dataset had a positive effect to stock returns over all the period and that this effect was due to mispricing. Using an event study analysis, Becchetti and Ciciretti (2011) found that companies with good ratings in the Product quality (or in the Governance) dimension reacted better to the Lehman Brothers' default. Since its introduction in 1995, Human rights seems to be the most discriminant item. However, this effect is probably due to the fact that most of the companies in our sample are not qualified for the Human rights KLD dimension (so a score 1 is rare on this item).

After 1997 Governance was not a very discriminant dimension, as it was very difficult to score 1 but also to score 0. Therefore the CSR ability is not strongly influenced by the Governance item whose effectiveness was widely criticized in literature (see for example Bhagat et al., 2008). For almost all years, Employment and Diversity were the easiest items both for 0 and 1 and they were not discriminant. This implies that the ability extracted from our model is less influenced by these items.

6 An application to portfolio allocations

For each year in the period 1992-2008, we ordered the companies according to the CSR ability η_p of the previous year. One of the more straightforward application of such ranking is stock selection. For benchmark to this ranking,

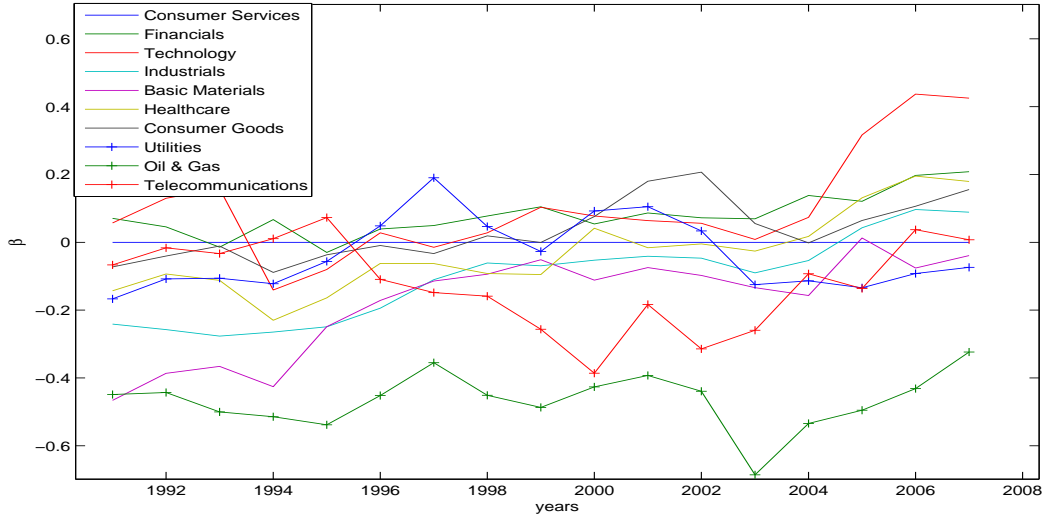


Figure 4: Industry effects. Evolution of the industry effects as measured by β_r coefficients, where r labels the industries

we ordered companies also according to the raw mean of their net score, denoted by η_p^0 , constructed by simply averaging the net score (3.1) over the different items.

We constructed high and low ranked portfolios according to both η_p and η_p^0 and compared their Jensen's α according to two different factor models: the first one taking the investment style into account and the second one measuring the exposure to industries.

6.1 Portfolio construction

Best(worst)-in-class portfolios are formed by taking, in each industry, companies above(below) the median value of η_p score. For benchmark reasons, we also construct best(worst)-in-class portfolios by taking companies, in each industry, according to the median value of η_p^0 . Figure 6 shows a scatterplot of η_p and η_p^0 with respect to their median values within each industry (year 2007). In order to strengthen the effect of the CSR information, we also considered, within each industry, the portfolios formed by picking the companies in the highest and lowest quartiles of the η_p or η_p^0 distributions. We denote

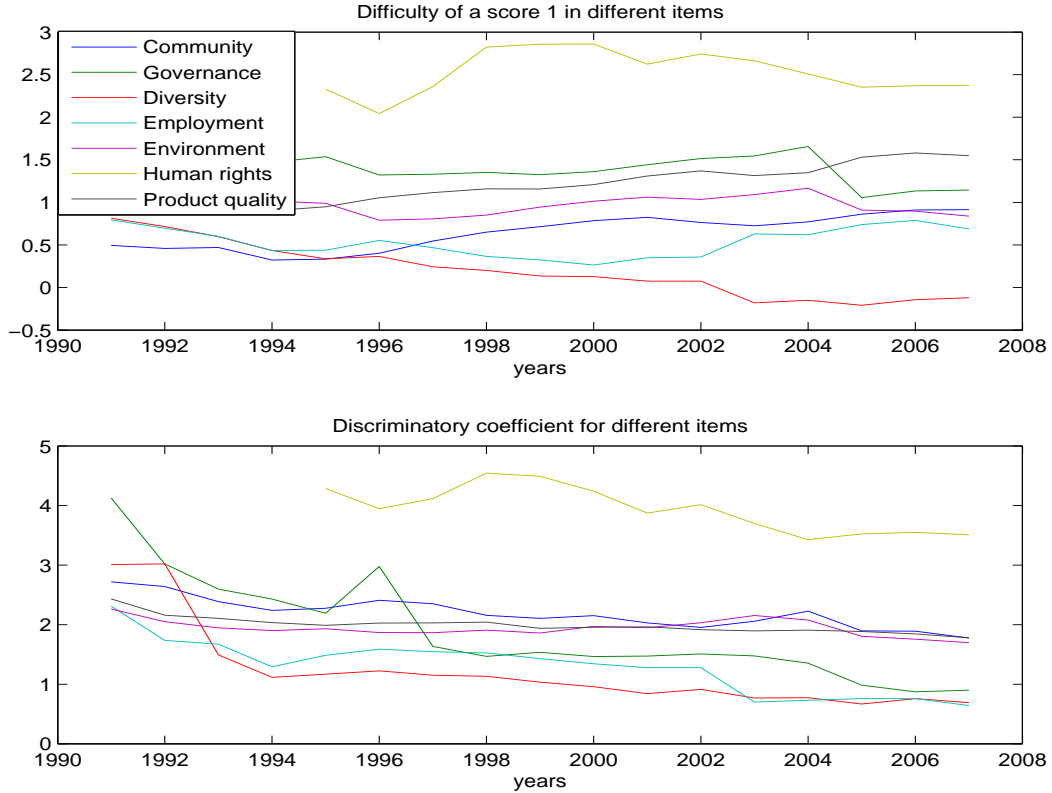


Figure 5: Model parameters. Evolution of the difficulty of $u = 1$ (upper panel) and of the discriminatory coefficient (lower panel) for all the items

with $L(L0)$ and $H(H0)$ respectively the high and low ranked portfolios according to $\eta_p(\eta_p^0)$. To keep the notation as simple as possible, we do not distinguish between portfolios formed on the basis of the median values from the ones formed on the basis of the quartiles, as it becomes clear from the discussion. Moreover, in order to make comparisons, we considered also the high minus low (i.e. $H0 - L0$, $H - L$, $H - H0$, $L - L0$) difference portfolios.

We formed three different types of portfolio: equally weighted (EW), value weighted (VW) and mean-variance optimal portfolios. Portfolio weights were re-balanced at the beginning of each year. Notice that while the EW and VW portfolios do not depend on any financial data, the optimal portfolios mix the ex-ante mean-variance optimality with the CSR information.

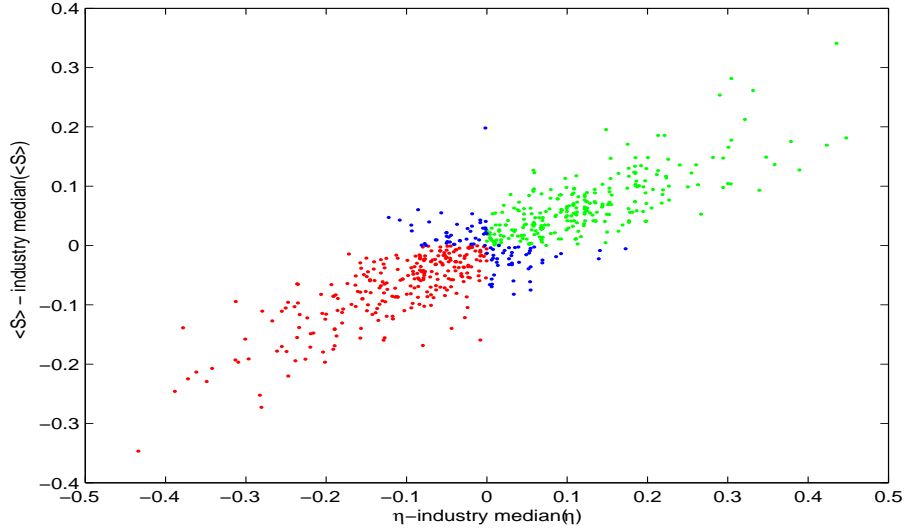


Figure 6: Comparison of the rankings. Scatterplot of $\eta(x)$ and $\eta^0(y)$ with respect to their median values within each industry (year 2007): in green(red) the companies that are high(low) ranked according to both the orderings. In blue those companies that are high ranked for one of the ordering but low ranked according to the other one

An optimal portfolio, in the mean-variance sense, is a portfolio that, given a certain level of expected return R , minimizes its variance keeping an expected return at least equal to R . Short-selling is not allowed in our analysis. Input data for optimal portfolio allocations, namely the vector of expected returns μ and the covariance matrix of the returns Σ , were estimated similarly to Herzel et al. (2011), using the Carhart's model (Carhart, 1997) to compute Σ and making a market neutral forecasting assumption for μ . In the construction of the optimal portfolios we set R as the level of the market expected return. That is an intermediate level that makes the optimal portfolios over the different subsets of allocation preserve a certain degree of diversification.

6.2 Measuring the portfolio risk-adjusted returns

We aim at measuring the managerial skills of different portfolio managers who use the extra non-financial information given by the CSR ability index η_p in their portfolio allocations. We also want to compare such performances with those of a manager who uses the more straightforward CSR index η_p^0 . For each portfolio strategy and their differences, we computed the intercept α with respect to two different multifactor models, covering different segments of the market.

The first model considered is the Carhart's model (Carhart, 1997) taking investment style into account. This is because, as observed for example in Kurtz (1997) and Guerard (1997), differences in investment style can be very relevant determinants in explaining the performance differences between ethical and conventional funds.

With $R_{j,t} - RF_t$ we denote the excess return over the risk-free rate RF_t of portfolio j at month t . The Carhart's model explains the portfolio returns in terms of 4 risk factors:

$$R_{j,t} - RF_t = \alpha_j + \beta_{j,1} (R_t^M - RF_t) + \beta_{j,2} SMB_t + \beta_{j,3} HML_t + \beta_{j,4} MOM_t + \epsilon_{j,t}$$

where $R_t^M - RF_t$ is the excess return of the market at time t , SMB_t is the return at time t of the small cap portfolio minus the large cap portfolio, HML_t is the return at time t of the value stocks' portfolio minus the growth stocks' portfolio, and MOM_t is the return at time t corresponding to the momentum factor⁵; the terms $\epsilon_{j,t}$ are the idiosyncratic errors and the β 's are the factor loadings of portfolios over the risk factors.

The second model implemented, similarly to Huij and Derwall (2011), is an industry model

$$R_{j,t} - RF_t = \alpha_j + \sum_r \beta_{j,r} INDUSTRY_{r,t} + \epsilon_{j,t}$$

where $INDUSTRY_{r,t}$ is the excess return of industry⁶ r at time t for r as in Section 4.

6.3 Portfolio performances

Results are summarized in Table 1 and in Table 2 respectively for the Carhart and the industry model. The tables refer to the case in which portfolios are

⁵The time series of the risk factors were downloaded from the K. R. French's web site.

⁶Proxies for the industry indexes were download from Datastream.

constructed picking the companies in the highest and the lowest quartiles. The tables show the ordinary least square estimates of the coefficients of the regressions together with their significance at levels 1%, 5% and 10%, indicated respectively with ***, ** and *. The levels of significance are computed taking a Newey-West correction into account. The Jensen's α 's are expressed as a percentage on a monthly basis.

First we comment on the intercept. The Jensen's α gives the portfolio extra return that cannot be obtained through the leverage to the risk factors and can be interpreted as a measure of the manager's picking ability after controlling for the factor effects. The α coefficients of the *EW* portfolios are not significant in the two models, and we will not discuss them. Table 1 shows that for *VW* and optimal portfolios the risk adjusted returns are positive and significant for high ranked portfolios according to the η_p ordering, while they are not significant for the low ranked ones. Moreover, the α 's of the *H* – *L* difference portfolios are positive and significant. It is remarkable that the risk-adjusted returns of portfolios based on η_p^0 (*H0*, *L0* or *H0* – *L0*) are never significant. The *H* – *H0* portfolios have α positive and significant. In the case of the optimal portfolio allocation also the *L* – *L0* portfolio has α significant but negative, a result that therefore reinforces the previous one.

Table 2 confirms the findings from the Carhart's model for *VW* and optimal portfolios. Although the α 's of high and low ranked portfolios are negative, independently from the ranking, again the high ranked portfolios according to the η_p ordering perform better than the low ranked ones, even though only in the case of the optimal portfolio in a significant way. On the other hand, the high ranked portfolios according to η_p^0 may perform even worse than the low ranked ones, although not significantly. Moreover we found that the α coefficients of the difference *H* – *H0* portfolios are positive and significant.

To summarize, according to both factor models, a portfolio manager investing in high ranked companies according to the CSR ability performs better than a manager who invests: (a) in the low ranked companies with respect to the same ordering; (b) in the high ranked companies according to the average of the scores.

Concerning the other factors in the Carhart's model, we notice from Table 1 that the market and the book to market ratio loadings of the high minus low portfolios are negative and significant in all the cases, according to the η_p ordering, and in most of the cases, according to η_p^0 . This means that the high ranked portfolios are usually less exposed to the market factor and are

invested more in growth stocks (low book to market ratio) than value stocks (high book to market) compared to the low ranked portfolios. These findings are in line with Bauer et al. (2005), where ethical and conventional funds from different geographic regions are compared.

The leverage of our portfolios on the industries are shown in Table 2. The high ranked *VW* and optimal portfolios are more exposed to the industry Technology and less to the industry Oil & Gas than the low ranked ones, while the high ranked *VW* portfolios are also less exposed to the industry Financials than the low ranked ones. The high ranked *EW* portfolios are instead more invested in the industries Healthcare and Oil & Gas and less in Basic Materials, Consumer Goods and Utilities than the low ranked ones. In any case, no particular pattern related to the industry exposure is observed.

Since results for portfolios formed by choosing firms with respect to the median values are in agreement with those here commented, we do not detail them here⁷. In general, when a coefficient is significantly different from zero for a portfolio built with respect to the median values, then that result (in terms of p-value or magnitude) is even strengthened in a portfolio built with respect to the quartiles. Moreover, in some cases we observed non significant coefficients of the median portfolios that became significant in the other case.

7 Conclusions

Corporate Social Responsibility has a multidimensional nature. Companies with a policy of responsibility have to comply with standards that are related to many different aspects. Industries can be highly relevant in determining the capacity of a firm to satisfy some requirements. We implemented a latent variable model that provides a univariate measure of the CSR performance of a firm, in such a way that the different aspects of CSR are weighted differently with the weights that are allowed to vary across industries. The model allows to rank firms according to their ability and items according to their discriminatory power.

We estimated the model on the KLD ratings of CSR of companies belonging to the S&P500 index and/or to the KLD 400 Domini Social index, covering the period from 1992 to 2008. The findings indicate that firms in the industry Oil & Gas have the highest difficulty to comply with the CSR

⁷Those results are available upon request.

standards, followed by the ones in Industrials, Basic Materials and Telecommunications. Environment, Community and Product quality are the most discriminant items. They provide a good measure of the firm's CSR ability to comply with CSR standards. On the contrary, Governance is not a discriminant item.

As an application, we ordered firms according to their CSR ability and performed portfolio selection. We found that the high ranked portfolios outperformed the low ranked ones in terms of the Jensen's α computed in (a) the Carhart's model (b) a factor model taking the industry exposure into account. As a benchmark, we ranked companies also on the basis of a more common and straightforward CSR index that is simply the average of the KLD net scores over the different CSR dimensions. High ranked portfolios according to the CSR ability outperformed the high ranked portfolios according to the row average of the CSR net scores. These findings suggest that the proposed synthetic measure of the CSR performance of firms may be used in socially responsible portfolio selection.

Acknowledgements

The authors gratefully acknowledge financial support from the Swedish Foundation for Strategic Environmental Research (MISTRA) through the Sustainable Investment Research Platform (SIRP).

References

- Barnett, M.L., Salomon, R.M., 2006. Beyond Dichotomy: the curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal* 27, 1101-1122.
- Bauer, R., Koedijk, K., Otten, R., 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking & Finance* 29, 1751-1767.
- Becchetti, L., Ciciretti, R., 2011. Stock Market Reaction to the Global Financial Crisis: testing for the Lehman Brothers' Event. *Giornale degli Economisti Italiani in the Special Issue of the Italian Econometric Society*, forthcoming.
- Benson, K.L., Brailsford, T.J., Humphrey, J.E., 2006. Do socially Responsible Fund Managers Really Invest Differently?. *Journal of Business Ethics* 65, 337-357.
- Bhagat, S., Bolton, B., Romano, R., 2008. The promise and peril of corporate governance indices *Columbia law review* 108, 1803-1882.
- Brammer, S., Brooks, C., Pavelin, S., 2006. Corporate social performance and stock returns: UK evidence from disaggregated measures. *Financial Management* 35, 97-116.
- Carhart, M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57-82.
- De Boeck, P., Wilson, M., 2003. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York, NY.
- Derwall, J., Guenster, N., Bauer, R., Koedijk, K., 2005. The Eco-Efficiency Premium Puzzle. *Financial Analyst Journal* 61, 51-63.
- Derwall, J., Koedijk, K., Horst, J.T., 2011. A tale of Values-Driven and Profit-Seeking Social Investors. *Journal of Banking and Finance* 35, 2137-2147.
- Guerard, J.B., 1997. Is there a cost to being socially responsible in investing?. *The Journal of Investing* 6, 11-18.
- Herzel, S., Nicolosi, M., Stărică, C., 2011. The Cost of Sustainability in Optimal Portfolio Decisions. *The European Journal of Finance*, forthcoming.
- Hopkins, M., 2005. Measurement of corporate social responsibility. *International Journal of Management and Decision Making* 6, 213-231.

- Huij, J., Derwall, J., 2011. Global equity fund performance, portfolio concentration, and the fundamental law of active management. *Journal of Banking & Finance* 35, 155-165.
- Kempf, A., Osthoff, P., 2007. The Effect of Socially Responsible Investing on Portfolio Performance. *European Financial Management* 13, 908-922.
- Kurtz, L., 1997. No effect, or no net effects? Studies on socially responsible investing. *The Journal of Investing* 6, 37-49.
- Manescu, C., 2011. Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk?. *Sustainable Development* 19, 95-118.
- Skrondal, A., Rabe-Heskett, S., 2004. Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Boca Raton, FL.
- Statman, M., Glushkov, D., 2009. The Wages of Social Responsibility. *Financial Analysts Journal* 65, 33-46.

Carhart's model estimate						
	$\alpha(\%)$	Mkt	SMB	HML	Mom	R_{adj}^2
EW						
<i>H0</i>	0.13	0.89***	0.17***	0.38***	-0.18***	0.79
<i>L0</i>	0.10	1.00***	0.18***	0.45***	-0.26***	0.84
<i>H0 - L0</i>	0.04	-0.12***	0.00	-0.07	0.08***	0.24
<i>H</i>	0.15	0.88***	0.19***	0.40***	-0.17***	0.80
<i>L</i>	0.08	1.01***	0.18***	0.46***	-0.28***	0.84
<i>H - L</i>	0.07	-0.13***	0.01	-0.07*	0.10***	0.31
<i>H - H0</i>	0.01	-0.01	0.02*	0.01	0.01	0.03
<i>L - L0</i>	-0.02	0.01	0.00	0.01	-0.02**	0.04
VW						
<i>H0</i>	0.18	0.86***	-0.07	-0.18***	-0.17***	0.80
<i>L0</i>	0.05	0.88***	-0.16***	0.17***	-0.08***	0.80
<i>H0 - L0</i>	0.14	-0.03	0.09*	-0.35***	-0.09***	0.30
<i>H</i>	0.30**	0.80***	-0.07*	-0.17***	-0.16***	0.81
<i>L</i>	0.01	0.88***	-0.15***	0.19***	-0.07***	0.79
<i>H - L</i>	0.29**	-0.08**	0.08	-0.36***	-0.08***	0.31
<i>H - H0</i>	0.11*	-0.05***	0.00	0.01	0.01	0.07
<i>L - L0</i>	-0.04	0.00	0.01	0.01	0.01	0.01
Optimal						
<i>H0</i>	0.14	0.85***	0.01	0.15***	-0.19***	0.81
<i>L0</i>	0.12	0.90***	-0.05	0.21***	-0.20***	0.81
<i>H0 - L0</i>	0.03	-0.05*	0.06	-0.06*	0.01	0.05
<i>H</i>	0.22**	0.82***	0.03	0.15***	-0.18***	0.81
<i>L</i>	0.05	0.91***	-0.04	0.24***	-0.21***	0.82
<i>H - L</i>	0.17*	-0.09***	0.07**	-0.09***	0.03	0.13
<i>H - H0</i>	0.08*	-0.03*	0.02	0.00	0.00	0.03
<i>L - L0</i>	-0.06*	0.01	0.01	0.03**	-0.01	0.01

Table 1: Carhart's model estimate for equally weighted (*EW*) value weighted (*VW*) and optimal portfolios from January 1992 to December 2008. The table shows the OLS estimates of the coefficients. The Jensen's α is expressed as a percentage on a monthly basis. The significance at the 1%, 5% or 10% level is indicated respectively with ***, ** and *. The p-values, not shown in the table, are computed by taking the Newey-West correction into account. The rows labeled with *H* and *L* report the results for portfolios constructed respectively over the highest and lowest ranked companies (upper and lowest quartiles for each industries) according to the CSR ability. The rows labeled with *H0* and *L0* show the results for portfolios consisting of the highest and lowest ranked companies according to the mean of the CSR net score over the CSR dimensions. The table shows also the results for some difference portfolios.

Industry model estimate												
$\alpha(\%)$	Basic Materials	Consumer Goods	Consumer Services	Financials	Healthcare	Industrials	Oil & Gas	Technology	Telecommunications	Utilities	$F_{a,dj}^2$	
EW												
$H0$	0.03	0.15***	0.09*	0.17***	0.20***	0.10**	0.04	0.10***	-0.07**	0.05	0.83	
$L0$	-0.03	0.20***	0.17***	0.16***	0.18***	0.04	0.06	0.10***	-0.05	0.11**	0.84	
$H0 - L0$	0.06	-0.06**	-0.08***	0.01	0.02	-0.02	-0.02	0.04**	-0.02	-0.06**	0.24	
H	0.06	0.14**	0.08	0.17***	0.10*	0.06	0.07**	0.10***	-0.09**	0.08*	0.82	
L	-0.05	0.22***	0.17***	0.18***	0.18***	0.03	0.05	0.11***	-0.06*	0.12**	0.84	
$H - L$	0.12	-0.08***	-0.09***	-0.02	0.07**	0.00	0.04**	-0.01	-0.02	-0.04	0.30	
$H - H0$	0.03	0.00	-0.01	0.00	0.00	0.01	-0.01	0.00	-0.02**	0.03***	0.08	
$L - L0$	-0.03	0.02*	0.00	0.02	0.00	-0.01	-0.01	0.01	-0.02**	0.01	0.04	
VW												
$H0$	-0.27**	0.00	0.09**	0.04	0.16***	0.15***	0.10	0.31***	0.00	-0.01	0.86	
$L0$	-0.16**	0.01	0.11***	0.07*	0.22***	0.13***	0.16***	0.05***	0.05**	0.00	0.90	
$H0 - L0$	-0.11	-0.02	-0.02	-0.02	-0.06*	0.02	-0.07	-0.13***	-0.05	-0.02	0.49	
H	0.03	0.03	0.07	0.05	0.12***	0.17***	0.05	0.29***	0.02	0.00	0.86	
L	-0.17**	0.04	0.10**	0.07*	0.23***	0.12***	0.14***	0.06***	0.04	-0.02	0.89	
$H - L$	0.05	-0.01	-0.03	-0.02	-0.11***	0.04	-0.09	-0.11***	-0.02	0.03	0.43	
$H - H0$	0.15**	0.03	-0.02	0.01	-0.04**	0.01	-0.02	-0.02	0.02	0.02	0.14	
$L - L0$	-0.01	0.03	-0.01	0.00	0.01	-0.01	-0.02	0.01	-0.01	-0.03*	0.03	
Optimal												
$H0$	-0.11	0.06	0.11***	0.07	0.23***	0.12***	0.06	0.16***	0.00	0.04	0.85	
$L0$	-0.15	0.06*	0.13***	0.08	0.22***	0.13***	0.15***	0.08***	0.02	0.03	0.86	
$H0 - L0$	0.05	0.00	-0.02	-0.01	0.01	-0.01	-0.09**	-0.05**	-0.02	-0.01	0.13	
H	-0.02	0.06	0.09**	0.08*	0.21***	0.11**	0.08	0.15***	-0.01	0.05	0.84	
L	-0.19*	0.09**	0.13***	0.10*	0.23***	0.11**	0.12***	0.09***	0.02	0.03	0.88	
$H - L$	0.17**	-0.03	-0.04	-0.01	-0.02	0.00	-0.04	0.06***	-0.03	0.02	0.16	
$H - H0$	0.09**	0.00	-0.02	0.01	-0.02	-0.01	0.01	-0.01	-0.01	0.01	0.03	
$L - L0$	-0.03	0.03**	0.00	0.01	0.00	-0.02*	-0.03**	0.01	0.00	0.01	0.04	

Table 2: Industry model estimate for equally weighted (EW) value weighted (VW) and optimal portfolios from January 1992 to December 2008. The table shows the OLS estimates of the coefficients. The Jensen's α is expressed as a percentage on a monthly basis. The significance at the 1%, 5% or 10% level is indicated respectively with ***, ** and *. The p-values, not shown in the table, are computed by taking the Newey-West correction into account. The rows labeled with H and L report the results for portfolios constructed respectively with the highest and lowest ranked companies (upper and lowest quartiles for each industry) according to the CSR ability. The rows labeled with $H0$ and $L0$ show the results for portfolios consisting of the highest and lowest ranked companies according to the mean of the CSR net score over the CSR dimensions. The table shows also the results for some difference portfolios.

QUADERNI DEL DIPARTIMENTO DI ECONOMIA, FINANZA E STATISTICA

Università degli Studi di Perugia

Quaderni pubblicati nel 2011

83	Gennaio 2011	Davide CASTELLANI Fabio PIERI	Foreign Investments and Productivity Evidence from European Regions
84	Febbraio 2011	Stefano HERZEL Marco NICOLOSI Cătălin STĂRICĂ	The cost of sustainability on optimal portfolio choices
85	Marzo 2011	Pierluigi GRASSELLI Cristina MONTESI	Politiche orientate al bene comune e politiche attive del lavoro
86	Marzo 2011	Massimo DE FELICE Franco MORICONI	Un'estensione stocastica del modello "Fisher-Lange"
87	Marzo 2011	Mirella DAMIANI Fabrizio POMPEI Andrea RICCI	Temporary job protection and productivity growth in EU economies
88	Aprile 2011	Gaetano MARTINO Paolo POLINORI	Productive process innovation as sequential adjustment of the hybrid governance structure: the case of the poultry sector
89	Aprile 2011	Stefano CASTRIOTA Marco DELMASTRO	Inside the black box of collective reputation
90	Maggio 2011	Roberto BASILE Luigi BENFRATELLO Davide CASTELLANI	Spatial clustering and nonlinearities in the location of multinational firms
91	Giugno 2011	Andrea RICCI Robert WALDMANN	Job Security and Training: the Case of Pareto Improving Firing Taxes
92	Giugno 2011	Francesca PIERRI Alberto BURCHI Elena STANGHELLINI	La capacità predittiva degli indicatori di bilancio: una verifica sulle aziende umbre
93	Luglio 2011	Francesco VENTURINI	Product variety, product quality, and evidence of Schumpeterian endogenous growth: a note
94	Settembre 2011	Francesco VENTURINI	Looking into the black box of Schumpeterian Growth Theories: an empirical assessment of R&D races
95	Ottobre 2011	Massimo MANCINI	Labor effort and temporary jobs: evidence for Italy
96	Ottobre 2011	Marco NICOLOSI Stefano GRASSI Elena STANGHELLINI	How to measure Corporate Social Responsibility