UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 376

Stock returns in relation to environmental, social, and governance performance: Mispricing or compensation for risk?

Cristiana Manescu

June 2010

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se info@handels.gu.se

Stock returns in relation to environmental, social, and governance performance: Mispricing or compensation for risk? [a]

Cristiana Mănescu [b]

Abstract

Using detailed data on seven environmental, social, and governance (ESG) attributes for a long panel of large publicly-traded U.S. firms during July 1992-June 2008, only community relations were found to have had a positive effect on risk-adjusted stock returns, which effect was due to mispricing. Additionally, a changing effect of employee relations was found from positive during July 1992-June 2003 to negative during July 2003-June 2008. The positive effect was due to mispricing, but there is some evidence that the negative effect was compensation for low non-sustainability risk. A negative effect of human-rights and product-safety indicators on risk-adjusted stock returns in the more recent period was also found to be due to mispricing. The implications are that certain ESG attributes are value relevant but they are not efficiently incorporated into stock prices.

JEL: G12, M14, G14.

Keywords: SRI, sustainability, risk-factor test, market efficiency

[a]*This research was conducted as part of the research programme Behavioral Impediments to Sustainable Investment, funded by the Swedish Foundation for Strategic Environmental Research, MISTRA.*

[b]Centre for Finance, School of Business, Economics and Law, University of Gothenburg, Box 600, SE-40530, Gothenburg, Sweden. Tel.: +46-317864412. Fax: +46-317861326. Email: cristiana.manescu@cff.gu.se I am grateful to Bert Scholtens (Editor) and an anonymous referee, to my supervisors Cătălin Stărica and Martin Holmen, and also to Leonardo Becchetti (Discussant Licentiate Seminar), Evert Carlsson, Stefano Herzel, Lennart Hjalmarsson, Gordon Clark, Nils Kok, Eric Knight, Stephanie Giamporcaro, Constantin Belu, Rick Wicks, and Andreea Mitrut for valued input; as well as to seminar participants: at the 15th Annual International Sustainable Development Research Conference; at the Mistra Workshop Gothenburg; at the Oikos PRI Young Scholars Finance Academy; at the XVII International Tor Vergata Conference on Banking and Finance; at the GU Department of Economics; and at GU's Centre for Finance. I appreciate the use of financial data from Keneth French's website. Special thanks to Lisa Symonds of KLD Research & Analytics for quick answers regarding data questions. The usual disclaimer applies.

1. Introduction

Incorporating environmental, social, or governance (ESG) concerns into investment practices - so-called socially responsible investing - is increasingly popular (SIF, 2007). One of the major forces behind the surge in such investment is increasing demand for it from institutional and even individual investors, partly because of increasing awareness of environmental risk (e.g., climate change) and social risk (e.g., the crisis in Sudan) (SIF, 2007, :iv).

The trade-off between ESG performance and investment returns is difficult to analyze, both theoretically and empirically, primarily because of the multi-dimensionality of the ESG concept. On the one hand, most empirical evidence suggests that ”good” stocks, i.e., with high ESG scores, earn positive abnormal returns (Derwall et al., 2005; Statman and Glushkov, 2009). This is claimed to be due either to investors underestimating the benefits of ESG or overestimating its costs, i.e., mispricing the value relevance of ESG concerns, or to compensation for risk (Derwall et al., 2005). At the same time, few studies show that some good stocks earn negative abnormal returns, also explained as either mispricing or compensation for risk (Derwall and Verwijmeren, 2007).

On the other hand, there is strong evidence that ”shunned” stocks, i.e., of companies in the alcohol, gambling, tobacco, firearms, military, and nuclear industries, also earn positive abnormal returns (Statman and Glushkov, 2009; Hong and Kacperczyk, 2009). This is explained as the effect of social norms as norm-constrained investors discriminate against those companies, thus producing a ”neglect” premium in their risk-adjusted returns (Hong and Kacperczyk, 2009). A theoretical model developed by Heinkel et al. (2001) shows that, with a sufficient number of norm-constrained investors the lack of risk-sharing opportunities in the market will lead to higher cost of equity capital (i.e., higher expected returns) for shunned stocks (i.e., polluting firms in their model) and a lower cost of equity capital for acceptable stocks (i.e., non-polluting firms in their model).

The controversy about what relative risk-adjusted returns should be expected from ESG investments arises because of the non-exclusiveness of the two stock universes. ”Good” firms are generally those with an outstanding record with respect to at least one of the several ESG concerns. At the same time, ”shunned” firms are not necessarily those with the lowest record with respect to actual ESG scores, but are rather ignored by ethical investors simply because of the industry they are in, i.e., because of ethical beliefs.

This study analyzes whether it is mispricing or compensation for risk that underlies the positive or negative abnormal returns earned by firms with high ESG scores. First we will evaluate if there is an association between those ESG scores and the cross-section of stock returns, while controlling for other factors known to explain stock returns, such as beta, size, value, and momentum. Then, for those ESG concerns for which we find an effect we will implement the Charoenrook and Conrad (2005) test of whether the candidate ESG concern is a risk factor. If the required conditions for the test are not met, we can conclude that the effect observed is due to mispricing.

The ESG scores used here are based on the Kinder, Lydenberg, and Domini (KLD) Research and Analytics ratings of U.S. firms (1991-2006), the longest existing dataset on ESG concerns including community relations, corporate governance, diversity, employee relations, environment, product safety, and human rights dimensions. Two alternative aggregation techniques are used, together with a Best-in-Class method, which is helpful in addressing industry variation in the ESG scores. As large companies are alleged to disclose more ESG information, which is thus readily available for investors' decision-making, the focus here is on members of either the Standard and Poor's 500 index (S&P500) or the Domini Social 400 index (DS400). Using the members of these two indexes also minimizes the presence of companies for which a score of zero is likely to indicate lack of rating as opposed to neutral ESG performance.

Fama and MacBeth (1973) month-by-month cross-sectional regressions of monthly stock returns on the beta, size, book-to-market, and momentum risk-factors, as well as on seven ESG concerns (community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety) found that only the community relations indicator had a positive impact on stock returns. The risk-factor test shows, however, that the higher returns were not compensation for risk, but rather reflect mispricing.

In a further analysis on subperiods July 1992-June 2003 and July 2003-June 2008, we will find that there is a shift in the effect of some ESG concerns, consistent across several aggregations of ESG scores. During July 1992-June 2003, community relations and employee relations had a positive effect on stock returns, while during July 2003-June 2008 employee relations, human rights, and product safety had a negative effect. The risk-factor test finds some evidence of employee relations as a risk factor in the second period, but not for any of the others. Thus, firms with poor employee relations seem to carry a premium for high non-sustainability risk. The evidence is weak as only the necessary condition for employee relations is met, but not the sufficient condition. One explanation for the shift in the effect of employee relations is that investors may now more efficiently value good employee relations in the recent period, following an increase in the availability of such information.

The remainder of the paper is organized as follows: section 2 discusses the theoretical framework on the expected returns of socially responsible investments and sets the working hypotheses; section 3 presents the empirical strategy; section 4 describes how the ESG scores and the financial variables are obtained as well as how the risk-factor mimicking portfolios are constructed; section 5 presents and discusses the empirical results, and section 6 summarizes and draws conclusions.

2. Theoretical framework and hypotheses

Socially responsible investing (SRI) emerged in the U.S. in the 1960s, some decades later in the UK, Canada, and Australia (Kinder, 2005). Because it mainly involved aligning investors' portfolios with their values, it has been called **values-based SRI** (Kinder, 2005). By the late 1990s it had also developed another name, **value-seeking SRI**, which means identifying and incorporating into the investment process social and environmental criteria which might affect future financial performance (Kinder, 2005). Thus, SRI is an evolving notion and it is still unclear *why* investors would incorporate *any* of the ESG concerns in their investment process, as well as *which* ESG concerns they might prioritize.

Academics have hypothesized various explanations of high-ESG firms' expected returns relative to those of conventional firms mainly based on two arguments, one economic and one based on discriminatory tastes (Statman, 2006). The *economic argument* is that there are costs and benefits associated with ESG concerns, but it is not clear which are higher. Moreover, in order for costs and benefits to be efficiently reflected in stock prices, there must be sufficient information available on ESG performance for the market to incorporate it efficiently. According to *the discriminatory-tastes argument*, however, the relation between costs and benefits of ESG is secondary. Instead there might be enough investors who derive non-financial utility from ESG investing that they affect pricing irrespective of whether ESG is net costly or beneficial. Furthermore, there is a third argument which has attracted less attention, a *non-sustainability risk* argument. ESG performance might in fact affect the risk profile of firms by adding a non-sustainability risk component in addition to the market risk, size, book-to-market and other systemic risks documented theoretically and empirically.

Putting these arguments together, and their implications for stock returns, leads to three mutually-exclusive scenarios regarding the risk-adjusted returns of high-ESG firms relative to low-ESG firms, which are briefly discussed below. Renneboog et al. (2008) provide a critical review of the literature on SRI and thoroughly discuss the causes and the shareholder-value impacts of ESG, among other related issues.

"The no-effect scenario" is that there is no difference in the returns, adjusted for common risk-factors, of high-ESG firms relative to low-ESG firms. This is entirely consistent with the efficient markets hypothesis if the ESG performance of firms provides no information relevant for pricing (Statman and Glushkov, 2009). Even if ESG performance provides information relevant for pricing, if this information is publicly available and fully incorporated into asset prices, then there should still be no difference in the risk-adjusted returns of ESG and non-ESG firms (Wall, 1995). In this case we cannot distinguish whether ESG costs are higher or lower than ESG benefits by looking solely at stock returns. When controlling for common risk-factors in studies on socially responsible (SR) fund performance versus non-SR fund performance, this scenario has usually been confirmed (Bauer et al., 2005).

"Mispricing scenario" is that ESG performance has an impact on firms' cash-flow streams (i.e., it is value relevant[1]) but without enough information available it is not efficiently reflected in stock prices. This translates into either higher or lower risk-adjusted returns for high-ESG firms depending on the net benefit of ESG (by the economic argument). If the benefits of ESG outweigh their costs, but investors, on average, consistently underestimate the benefits or overestimate the

[1] This is the *value relevance hypothesis* discussed in Derwall and Verwijmeren (2007)

costs, then the risk-adjusted returns of high-ESG firms would be higher than of low-ESG firms (Statman, 2006). Underestimating the benefits of ESG could be reflected in positive earnings surprises (Edmans, 2008) or in reduced earnings volatility (Derwall and Verwijmeren, 2007), either of which could lead to mispricing. High performance with respect to the environment during 1995-2003 (Derwall et al., 2005) or employee relations during 1984-2005 (Edmans, 2008) have been found to provide positive abnormal returns, interpreted as mispricing. In a sample of multinationals, Dowell et al. (2000) also found that firms with high environmental standards had higher firm value than others, measured by Tobin's Q. Portfolios built on specific ESG dimensions have even been shown to have positive abnormal returns over long periods (Kempf and Osthoff, 2007).

On the other hand, it could be that, by the same economic argument, the risk-adjusted returns of high-ESG firms will be lower if the benefits of ESG are lower than its costs and incompletely informed investors overestimate benefits or underestimate costs. Barnea and Rubin (2006) show that ESG performance can be a source of agency costs because firms' managers have an incentive to promote ESG investment at the expense of shareholders' benefits in order to gain reputational benefits. Their hypothesis has empirical support as they found a negative link between insiders' ownership and the social rating of firms.

"Risk-factor scenario" is that expected returns of low-ESG firms are higher primarily because they carry a premium for non-sustainability risk. The ESG rating of a company might indicate its exposure to a non-sustainability risk factor. Besides environmental risk, this risk factor might include product and commercial-practices risks, or risk associated with workplace quality of life (Dufresne and Savaria, 2004). It can also relate to litigation risk, investor trust, and other intangible advantages (Becchetti and Ciciretti, 2006; Derwall and Verwijmeren, 2007) that might dramatically affect firms' future financial performance (i.e., value-seeking SRI). Given the increased awareness of sustainability risk, it is expected that the non-sustainability premium has increased in recent years.

A systematic distortion in market pricing resulting in higher expected returns for low-ESG firms (or non-SR firms) can *also* occur under the discriminatory-tastes argument that divestment from certain stocks is done exclusively on moral or ethical grounds (Hong and Kacperczyk, 2009; Derwall and Verwijmeren, 2007). The existence of investors who derive non-financial utility from investing with high-ESG firms (i.e., values-based SRI), could decrease the demand for the stock of low-ESG firms and thus increase their cost of equity capital. Heinkel et al. (2001) developed an equilibrium model in which there is an impact on firms' costs of capital as soon as there is a substantial share of SR investors.[2] Though expressed in environmental terms, the model could easily be extended to social and governance concerns. Hong and Kacperczyk (2009) have shown

[2]This model belongs to the family of theoretical models explaining empirical pricing anomalies (e.g., the equity-size premium anomaly) through differences in information quality or quantity (asymmetric information) across investors (Arbel et al., 1983; Merton, 1987). These models show that in equilibrium it is both market risk and idiosyncratic risk that affect expected returns, due to distortions induced by different investor bases.

that higher institutional divestment from "sin stocks" leads to higher cost of capital for these firms relative to others, confirming this alternative argument.

Alternatively, under the risk-factor scenario, it can also be that expected returns of high-ESG firms are higher because they might carry a premium for some missing risk-factors, others than the common factors beta, size, value and momentum. This argument has usually been used as alternative explanation to mispricing for the higher risk-adjusted returns observed for high-ESG firms, such as the eco-efficiency premium puzzle (Derwall et al., 2005).

Either because of ethical beliefs (values) or a (non-)sustainability risk-factor, ESG performance might thus affect expected stock returns, if two conditions are met: Information on ESG performance must be available to investors, and there must be enough investors who care. Since both of these can change, the effect of ESG performance on stock returns can vary over time.

Regarding availability of ESG information, the UK Companies Act of 2006 holds company directors responsible for disclosing environmental and social information regarding firms' long-term prospects (Clark and Knight, 2009). According to CorporateRegister.com, a private company specializing in tracking corporate social responsibility/ESG disclosures, since the mid-1990s more and more firms have started disclosing ESG information (Dhaliwal et al., 2009). Dhaliwal et al. argue that the increasing voluntary ESG disclosures might have to do with increased scrutiny of the corporate impact on society following the loss of trust after the series of corporate scandals in the early 2000s. Finally, KPMG (2008) reports that 80% of the 2200 largest corporations worldwide published an ESG report in 2008 or integrated ESG information into their annual reports.

Regarding the number of socially responsible investors, Haigh and Hazelton (2004) question the existence of a large number, while Hong and Kacperczyk (2009) argue the opposite. Epstein and Freedman (1994) surveyed a random sample of average individual investors and found that they also wanted ESG information, not just institutional investors. Those surveyed primarily wanted information on environmental performance and product quality. Empirical studies also show market reaction to disclosure of such information (Patten, 1990), confirming that investors care.

The working hypothesis here is that, during 1992-2008, there was a positive association between candidate ESG variables and risk-adjusted stock returns, due to ESG benefits outweighting costs, but without enough ESG information available. This hypothesis will be confirmed if the Charoenrook and Conrad (2005) test rejects the null that any particular ESG variable is a risk-factor. Moreover, given the rise in sustainability-risk awareness and in information availability in recent years, we can also expect that, for subperiods, some of the ESG variables might function as systematic risk-factors (i.e., non-sustainability risk-factors). Therefore, the second working hypothesis here is that in recent years we might find a negative association between some ESG variables and risk-adjusted stock returns which could be due to compensation for lower sustainability risk. Again, this will be tested by the Charoenrook and Conrad (2005) test.

Many studies have pointed towards such positive or negative associations but without explaining what lies behind them. Edmans (2008) is among the very few who tested for mispricing (by looking at positive earnings surprises) as explanation for the positive association between employee relations and risk-adjusted stock returns.

This study aims at narrowing this gap by explicitly testing whether any risk-adjusted returns reaction to ESG performance is compensation for risk or rather is due to mispricing. It may also help to better understand the puzzle that one observes both high returns for high-ESG firms and high returns for ethically excluded firms, formulated by Statman and Glushkov (2009).

3. The Empirical Strategy

The central aim of this paper is to test whether the explanatory power that some ESG concerns have for stock returns is due to mispricing or compensation for risk.

Explanatory power is assessed using cross-sectional regression of excess stock returns on ESG scores and four factors known to explain stock returns: beta, size, value, and momentum, i.e., the Fama-French three-factor model (Fama and French, 1992) plus the momentum factor identified by Jegadeesh and Titman (1993). The cross-sectional approach is preferred to the portfolio approach of analyzing only the return differential between high- and low-ESG stock portfolios because of interest in the monotonic effect of ESG concerns on stock returns. Moreover, the use of ex-post returns should supplement Derwall and Verwijmeren (2007), who used measures of ex-ante returns, i.e., implied cost of equity capital.

The test for the effect of ESG concerns is whether their estimated effect is statistically indistinguishable from zero as the four-factor model predicts. Alternatively, if ESG concerns are relevant for explaining variation in returns, we can implement the Charoenrook and Conrad (2005) test of whether a candidate variable is a risk factor or not. The test is whether there is a relationship between the conditional mean and the conditional variance of the return of the candidate variable's factor-mimicking portfolio, and whether risk, in the mean-variance framework, explains all the return of a mimicking portfolio.

The central economic question that the cross-section regressions can answer is why average returns vary across assets (Cochrane, 2005). The expected return of an asset should be high if it has large exposure to factors that carry risk premiums, i.e., market risk or beta, size (market capitalization), book-to-market ratio, or momentum (a simple average of 10 past returns).

The four-factor model estimated here is extended with several ESG variables - or alternatively with an aggregate ESG variable - as

$$R_{jt+1} = \gamma_0^{t+1} + \gamma_1^{t+1}\widehat{\beta_{jt}} + \gamma_2^{t+1} Size_{jt} + \gamma_3^{t+1} BookToMarket_{jt} + \gamma_4^{t+1} Momentum_{jt} + \gamma_5^{t+1} ESG_{jt} + \theta_{jt+1} \quad (1)$$

where the excess stock return for firm j in month $t+1$ (R_{jt+1}) is a function of $\widehat{\beta_{jt}}$, the estimated market risk (beta) of the firm; $Size_{jt}$, the firm's log of market capitalization; $BookToMarket_{jt}$, its book-to-market ratio; $Momentum_{jt}$, the average return over the period $t-2$ and $t-12$

months; ESG_{jt}, seven individual firm ESG variables or an overall ESG variable; and an i.i.d. error term, θ_{jt}, with zero mean and constant variance. Size and book-to-market ratio are updated monthly (as did Galema et al., 2008), while estimated beta and the ESG variable(s) are updated every year. Beta is estimated for each asset j ($j = 1...N$) through a time-series regression up to time t of the asset's returns and the market-index return.

Empirical evidence shows that not all ESG concerns are equally relevant to investors or, perhaps more importantly, that there are confounding effects between some of them. Thus the focus on the individual ESG concerns in the analysis, also complemented with an aggregate ESG score.

This model (1) was estimated with the Fama-MacBeth procedure, which allows for time-varying coefficients ($\widehat{\gamma_k^{t+1}}$, $k = 1, 2, 3, 4$). This is known as a *two-pass* Fama-MacBeth estimation, because one first estimates a time-series regression of individual stock returns on the market-index return in order to obtain beta estimates, and then uses these estimates as explanatory variables in the cross-sectional regression (1). Estimation bias can be a problem when using this procedure, due to possible measurement error (sampling variance) in $\widehat{\beta_j}$, i.e., the errors-in-variables (EIV) problem (Black et al., 1972). The estimated $\widehat{\beta_j}$ in (1) equals the true (unobservable) value plus a measurement error (or sample variance) v_j, assumed to be i.i.d. with zero mean and variance σ_v^2. The "grouping technique" developed by Black et al. (1972) provides N-consistent estimates of $\widehat{\gamma_1}$ and was thus used in what follows (see detailed description of this approach in Appendix A).

Fama and MacBeth (1973) indicate that the mean values of the $\widehat{\gamma_k^{t+1}}$ coefficients, which are needed for statistical inference, can be computed as the time series averages of the estimated coefficients, i.e., $\bar{\gamma}_k = Avg(\widehat{\gamma_k}) = \sum_{t=1}^{T} \widehat{\gamma_k^t}/T$. Then the t-test is the ratio of this mean and the time-series standard error of the estimated coefficient, i.e., its time-series standard deviation ($sd(\widehat{\gamma_k})$) multiplied by the square root of the time-series length (T), $t(\hat{\gamma}_k) = \frac{\bar{\gamma}_k}{sd(\widehat{\gamma_k})/\sqrt{T}}$, $k = 0, 1, 2, 3, 4$).

Most empirical studies of ESG point out that ESG is industry-specific. For example, firms in a certain industry might have both high ESG scores and high stock returns, while those in another industry might have low ESG scores and low returns. Without controlling for such industry effects, a false positive association might thus appear between ESG and returns. Conversely, any of the ESG concerns that might have different effects across industries would obscure their overall effect. Therefore, Equation (1) was also estimated augmented with 9 industry dummies Ind_i, to control for any confounding effects, as

$$R_{jt+1} = \gamma_0^{t+1} + \Sigma_{i=1}^{9} \alpha_i^{t+1} Ind_i + \gamma_1^{t+1} \widehat{\beta_{jt}} + \gamma_2^{t+1} Size_{jt} + \gamma_3^{t+1} BookToMarket_{jt} + \\ + \gamma_4^{t+1} Momentum_{jt} + \gamma_5^{t+1} ESG_{jt} + \theta_{jt+1} \tag{2}$$

which is the main estimation model.

The test for a candidate ESG variable being a risk factor was developed by Charoenrook and Conrad (2005) based on the relationship between the conditional mean and the conditional variance of return on the candidate variables's factor-mimicking portfolio. Specifically, with a linear factor structure, they note that the risk premium at time t on the factor-mimicking portfolio, like that of any asset, is related to the conditional variance of the portfolio's return with the pricing kernel. However, for a factor-mimicking portfolio, the conditional variance with the pricing kernel is linearly related to the factor's conditional variance, and hence to the portfolio's variance. Therefore, there must be a linear relationship between the conditional mean and the conditional variance of the factor-mimicking portfolio return, if that factor is a component of the pricing kernel. For a formal deduction of this relation see Section III in Charoenrook and Conrad (2005).

To test this relation, they propose a Garch-in-Mean family of estimations, among which the main one is

$$\begin{aligned} R_{X,t+1} &= \mu + \delta h_{t+1} + \eta_{t+1} \\ with \qquad h_{t+1} &= \omega + \alpha\eta_t^2 + \gamma h_t \end{aligned} \tag{3}$$

where $R_{X,t+1}$ is the excess return on the mimicking portfolio for factor X at time $t+1$; $h_{t+1} = Var_t(R_{X,t+1})$ is the expected conditional variance for the factor portfolio at time $t+1$; and $\eta_{t+1}/\sqrt{h_{t+1}}$ is normally distributed.

Three hypotheses were tested related to Equation (3). First, if X is a risk factor, then the relation between the conditional mean and variance of the portfolio return, estimated with coefficient δ in Equation (3), should have the same sign as the conditional expected risk premium on X (and be statistically significant).

Second, the intercept term (μ) in the conditional mean equation of (3) should be zero, meaning that the expected risk premium for the portfolio should be given entirely by the portfolio's conditional variance. A non-zero and statistically significant intercept might represent a component of the factor-mimicking portfolio unrelated to risk or an unmodeled time-variation in the price of (variance) risk.

Third, if X is a proxy for risk, then the Sharpe ratio of the factor-mimicking portfolio should be plausible, i.e., it should be less than that of the *ex ante* tangency portfolio. MacKinley (1995) argues that a reasonable value for this is approximately 0.17 per month, while Ahn et al. (2005) found a 0.39 Sharpe ratio from an efficient frontier built with 25 size- and book-to-market sorted portfolios for the 1959-2003 sample period.

Here, Equation (3) was estimated for mimicking portfolios for those ESG concerns found to explain excess returns in the cross-sectional approach. If the conditions above are not met, we have evidence that those ESG concerns' effects on stock returns were due to mispricing or irrational investor behavior. If all conditions are met, we have evidence that the ESG concerns functioned as risk-factors.

4. The Data

4.1. **The ESG Dataset.** Measuring ESG concerns is not easy, and measuring it over a long period as required for risk-factor analysis is especially difficult. Therefore the expertise of the investment research firm Kinder, Lydenberg, and Domini (KLD) Research & Analytics was used. They are the leading authority on social research and indexes for institutional investors (their database is preferred in many empirical analyses on SRI). Since 1991 they have collected data about seven ESG dimensions: community relations, corporate governance (labeled Other until 2002), diversity, employee relations, environment, human rights, and product safety; for 650 publicly-traded U.S. firms (members of either the S&P 500 or the DS400 index) until 2000, then 1100 firms during 2001-2002, and about 3100 (members of the Russell 3000 index) since 2003. The data comes from media articles, company documents, and direct communication with company officers, as well as government and NGO information. The research process, however, is proprietary. The previous year's annual ESG dataset is available for purchase at the end of each January or early February.

Each ESG dimension is graded annually based on a number of positive and negative indicators, i.e., strengths and weaknesses, which are given a score of 1 if present, otherwise 0. As KLD data has been used extensively in the ESG literature, the complete list of indicators can be found in Appendix B (for a thorough description of the data see Becchetti and Ciciretti, 2006; Derwall and Verwijmeren, 2007). The total number of indicators has varied from 54 in 1991 to 72 in 2006. The strengths and weaknesses for each ESG dimension are then aggregated to give an overall score for that dimension.

The most popular aggregation method, used in most studies based on KLD data (e.g., Derwall and Verwijmeren, 2007; Chatterji et al., 2008), has been to take the sum of strengths net of the sum of weaknesses, i.e.,

$$ESG_t^j = \sum_{s=1}^{u_t^j} strength_s^j - \sum_{r=1}^{k_t^j} weakness_r^j \tag{4}$$

where ESG_t^j = ESG dimension j, year t; u_t^j = number of strengths for dimension j, year t; $strength_s^j$ = strength indicator, equal to 1 if the firm meets strength s, j, otherwise 0; k_t^j = number of weaknesses for dimension j, year t; $weakness_r^j$ = weakness indicator, equal to 1 if the firm meets weakness r, j, otherwise 0.

The drawback with this approach is lack of comparability across years and across dimensions, as the number of strength and weakness indicators for most have varied. The product safety and environment dimensions have been quite stable since 1991, but the other dimensions have varied considerably.

An alternative aggregation method, used by Kempf and Osthoff (2007) (hereinafter the ”KO” method), involves first converting weaknesses into strengths by taking binary complements (meaning that if a certain weakness is not present, i.e., rated as 0, it is considered a strength rated with 1; if the weakness is present, then its corresponding strength is rated 0) and then summing up over all ”strengths” for a certain ESG concern. The results are then normalized to maximum for each dimension, by

$$ESG_t^j = \frac{\sum_{s=1}^{u_t^j} strength_s^j + \sum_{r=1}^{k_t^j} (!weakness_r^j)}{u_t^j + k_t^j} \tag{5}$$

where $!weakness_r^j$ = transformed strength indicator, equal to 1 if the firm does not meet weakness$_{r,j}$, otherwise 0; and ESG_t^j, u_t^j , k_t^j , and $strength_s^j$ are the same as in Equation (4). The scores thus range from 0 to 1.

This method is still subject to two criticisms concerning the representativeness and firm-level comparability of the results for at least five of the seven ESG dimensions.

The first criticism mainly concerns the environmental dimension, which is especially relevant in industries such as oil & gas, basic materials, industrials, and healthcare, while less relevant in others. Thus, for example, if the weakness ”ozone depleting chemicals” (as well as other weaknesses in the environmental dimension) are not present, this can certainly be seen as a strength for the oil & gas industry, but it is irrelevant for the telecom and financial industries. Without adjustment for this, firms in industries less affected by environmental strengths or weaknesses would receive inflated environment scores. Similarly a weakness indicator in the community dimension is ”investment controversies”, defined as ”a financial institution (!) that had lending or investment practices that lead to controversies”. Again, without adjustment, firms in other industries would receive inflated community scores.

The other criticism of the KO method is that not having certain weaknesses (such as having recently paid substantial fines for violations of employment standards) automatically becomes a ”health and safety” strength which competes with the already defined ”health and safety” strength (if the firm has made significant effort to improve employment conditions). However, not having paid fines for violations of employment standards is not the same as making effort for improvement. The firm has then complied with common rules, making it neutral with respect to the employee dimension, but not necessarily a leader. Examples in other ESG dimensions are ”South Africa” weakness in the human rights dimension, meaning that the company had faced controversies because of its operations in South Africa (which is irrelevant if the firm has no operations in S.A., yet automatically gets a ”South Africa” strength), or ”indigenous people relations” weakness in the community relations dimension.

Thus the KO method might *systematically* distort the aggregate scores for at least four of the ESG dimensions: community relations, employee relations, environment, and human rights.

As both aggregation methods, i.e., Equations (4) and (5), are flawed, a new one answering these criticisms (hereinafter called the "Relative" method) was constructed. It is based on Equation (4), but in addition standardizes the numbers of strengths and weaknesses in each dimension by the corresponding annual numbers of strengths and weaknesses (similar to the KO method). This avoids the two criticisms of the KO method and provides year-to-year comparability that was missing in the procedure in Equation (4). Formally, the method is

$$ESG_t^j = \frac{\sum_{s=1}^{u_t^j} strength_s^j}{u_t^j} - \frac{\sum_{r=1}^{k_t^j} weakness_r^j}{k_t^j} \tag{6}$$

By construction, ESG scores based on this method range from -1 to 1. The annual overall firm-ESG score across all seven dimensions is then simply their average ($\sum_{j=1}^{7} ESG_t^j/7$).

For comparison purposes, both the Relative and the KO method are used in the empirical analysis, but the Relative method is considered more suitable, not least because it is only a slight modification of the standard aggregation method used in almost all studies on KLD data.

In addition, a Best-in-Class score, by either the Relative or the KO method, is also used in order to embed industry differences already in the ESG scores. The Best-in-Class score is computed as the firm-score deviation relative to its industry mean score. This is a very frequent method in the ESG literature (e.g., Kempf and Osthoff, 2007; Statman and Glushkov, 2009). In the end, four measures for ESG performance are used: the Relative score, the KO score, the Relative Best-in-Class (Relative BC), and KO Best-in-Class (KO BC).

4.2. **The Financial Dataset.** The financial variables matched with the KLD data were obtained from Datastream. KLD data contains firm-year observations, identified since 1995 by their CUSIP code.[3] Missing CUSIPs before that were supplied by referencing firm-name and ticker identification with firms in 1995. Based on these CUSIPs, ISINs were next obtained, followed by Datastream codes, the unique security identification key in Datastream.[4] Datastream codes could be identified for 96.6% of the firms in the KLD dataset.

To ensure that accounting variables (in financial statements) were available before the stock returns they are used to explain, equity book-values for all fiscal year-ends in calendar year $t-1$ (1991-2006) were matched with returns for July of year t (1992-2007) to June of year $t+1$ (1993-2008). The 6-month (minimum) gap between fiscal year-end and return-period is conservative, similar to that used in Fama and French (1992).

Market equities at the end of December of year $t-1$ were used to compute book-to-market ratios, while logs of market equities at the end of June of year t gave firm size. Both variables were updated for each month. Therefore, to be included in the return tests, a firm must have had available the book value (Item 03480 in Datastream) as of December 31 of year $t-1$ and market

[3]CUSIP (Committee on Uniform Security Identification Procedures) is the North American security identification code.
[4]ISIN=International Securities Identifying Number.

value of equity (Item MV) as of June 30 of year t. Moreover, in order to obtain the pre-ranking beta estimates, monthly returns for at least 24 months preceding July of year t must have also been available.

Monthly stock returns were obtained based on the Datastream Item Return Index.[5] The excess stock-return series used in the empirical tests is the stock return less the risk-free rate. A proxy for the monthly risk-free rate, as well as for the return series of U.S. market index, of SMB and HML portfolios, were obtained from K. French's website.[6]

The most sensitive variable in the empirical tests is the firm β estimate, which was obtained by applying the portfolio grouping technique as in Fama and French (1992) with a minor modification due to different sample sizes, described in Appendix A.

4.3. **The Risk-Factor Mimicking Portfolios.** The non-sustainability risk-factor mimicking portfolios mirror the book-to-market mimicking portfolio used in Fama and French (1993). Firms were first sorted independently in two size (Small, Big), two book-to-market (Value, Growth), and three ESG-score (sustainability) portfolios, resulting in 12 portfolios. The size and book-to-market breakpoints were the median sample market-equity and median book-to-market, respectively. The sustainability breakpoints which determined the Low, Medium, and High Sustainability portfolios are the 30th and 70th sample percentiles. The returns on the portfolios were value-weighted. (Equally-weighted portfolios were also used without altering the final results). The Low-Sustainability minus High-Sustainability (LMH) risk factor was defined as

$$\begin{aligned} LMH &= 1/4(SmallValueLowSust. + SmallGrowthLowSust. + BigValueLowSust. \\ &+ BigGrowthLowSust.) - 1/4(SmallValueHighSust. + SmallGrowthHighSust. \\ &+ BigValueHighSust. + BigGrowthHighSust.) \end{aligned}$$

Portfolio returns were computed monthly, from July of year t to June of year $t+1$, and rebalanced every year based on market equities as of June of year t, book-to-market as of December of year $t-1$, and ESG scores available Jan-Feb of year t. Using a ranking on size and book-to-market minimizes the effects of these factors on the LMH risk-factor of interest. This is the non-sustainability risk-factor mimicking portfolio on which the Charoenrook and Conrad (2005) test was later applied.

5. Empirical Analysis

5.1. **Sample Description.** As noted, the sample consisted of members of the S&P500 and the DS400 index. The initial dataset consisted of an average of 639 firms per year for the 16 years

[5]The Item Return Index accounts for stock splits and dividends by adding back the dividend amount to the ex-dividend-date closing price.

[6]$http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html\#Benchmarks$, accessed September 2008.

from 1991 to 2006, with 10,220 firm-year observations, reduced to 10,174 that also had industry classification. Summary statistics for the seven individual ESG dimensions and an aggregate score, computed with the Relative and KO aggregation methods as well as their Best-in-Class equivalents, are presented in Table 1.

For both aggregation methods, higher score indicates better ESG performance. With respect to the Relative aggregation method (Table 1, Panel 1), a mean lower than 0 for corporate governance, environment, human rights and product safety dimensions indicates that the number of weaknesses outweighed the number of strengths. Identical values of 0 for the 25th-, 50th- and 75th-percentiles for environment, human rights and product safety dimensions show that 50% of firms recorded no strengths or weaknesses in these dimensions. As they are members of S&P500 or DS400, it is very unlikely that this is a result of KLD analysts not rating these firms at all.

While it is fairly easy to infer general characteristics of the ESG sample based on the Relative aggregation method, it is not straightforward to do so based on the KO method (Table 1, Panel 2). For example, a mean lower than 0.5 would indicate a low Relative ESG performance. However, by the KO measure, it might also indicate a much higher number of strengths than weaknesses, which imbalance tends to lower the aggregate scores. For example, diversity has a KO mean of 0.33, meaning low diversity performance at first, which is an artifact of the fact that there were 8 strengths and only 3 weaknesses measured. Standard deviations for most dimensions are about one-fifth of the corresponding mean, indicating low variation.

The Best-in-Class scores make the two aggregation methods directly comparable (Table 1, Panels 3 and 4). There are very similar distribution patterns of the two aggregation methods, except that both standard deviations and values of the Relative scores are double the size of the KO scores, which is achieved by construction. Estimated correlation coefficients between the ESG dimensions (not shown) are small, ranging from -0.15 to 0.20 regardless of method.

The requirement of minimum 24 monthly stock returns prior to July of year t (1992-2007) and equity book-value as of December 31 of year $t-1$ (1991-2006) reduced the sample to 9,371 firm-year observations. Eliminating outliers in monthly stock returns (over 60% a month) and in book-to-market ratios (negative and larger than 3) reduced the sample further to 9,202 firm-year observations, or 110,427 firm-month observations.

The logs of market equities (Table 2) indicate, as expected, that the sample consists of large firms, which is also indicated by an average beta lower than 1 and negative skewness in its distribution. The sample firms were classified into ten industry groups (sample distribution in parentheses): oil & gas (4.94%), basic materials (7.68%), industrials (18.95%), consumer services (14.02%), healthcare(7.39%), consumer goods (15.38%), telecom (2.09%), utilities (7.39%), financials (12.69%), and technology (9.47%).

5.2. **Analysis Results and Discussion.** Equations (1) and (2) were estimated using Fama-MacBeth month-by-month cross-sectional regressions, with the ESG variables computed with the Relative and KO methods along with their Best-in-Class counterparts. During 1992-2008, book-to-market and momentum explained the cross-section of returns while beta and size did not, irrespective of whether industry effects were controlled for or not (Tables 3 and 4). Moreover, as expected, book-to-market and momentum had positive effects on stock returns, the magnitudes of each, i.e., risk premiums, being influenced by the units of measurement of the underlying factor, so they must be interpreted with caution. For example, with industry dummies, a one-standard-deviation increase in book-to-market ratio and in momentum led, respectively, to 1.39-1.22% and 2.56% higher average annual returns, based on the marginal estimated effects.

With respect to the ESG variables, only community relations had an effect on stock returns, which was especially strong when industry effects were controlled for (Table 4). In this case community relations, computed by both Relative and KO methods as well as their Best-in-Class counterparts, had positive effect on stock returns, while without industry dummies (Table 3) a statistically significant effect is found only when using the Relative Best-in-Class method, which also controls to some extent for industry effects. Thus it is important to control for industry effects on ESG scores, and the Best-in-Class method appears to be the superior aggregation method. In terms of economic value, a one standard-deviation increase in the community relations score led to 0.52% (based on Relative scores) or 0.58% (based on KO scores) higher returns per year.

The aggregate ESG variable had no statistically significant effect on stock returns over this period, which is evidence of confounding effects between the ESG dimensions (since the community relations score was significant) and therefore they should be analyzed individually. This result is consistent with the findings in Derwall and Verwijmeren (2007) for the lack of effect of an aggregate ESG variable.

In the model including industry dummies (Table 4), it makes no difference if one uses Best-in-Class ESG scores or plain scores, as the estimation produces identical results for each aggregation method and its Best-in-Class counterpart (Columns 1 and 3, or Columns 2 and 4). Industry dummies thus seem to completely capture the between-industry variation in the ESG scores. In addition, the Relative and KO scores deliver a very similar pattern of estimates, except for the obvious differences in magnitude.

Next we investigate whether the positive effect of the community relations score was due to mispricing or whether it could be compensation for risk, with a premium for the firms with high community scores. For this reason a Low-Minus-High (LMH) community-risk-factor mimicking-portfolio was built following the procedure described in Section 4.3, using the Relative aggregation method and its Best-in-Class counterpart. To ensure that the High and Low community portfolios are mutual exclusive, the 20th- and 80th-percentile breakpoints were used with the Relative scores, and the 30th- and 70th-percentile breakpoints were used with their Best-in-Class counterpart.

Garch-in-Mean estimation of Equation (3) for the community-mimicking portfolio, using either the Relative or the Relative Best-in-Class method of obtaining the scores, shows that the conditional mean return was not linked to its conditional variance, as the δ estimate was not statistically different from zero, i.e., p-value$_\delta$ $= 0.32$ or 0.40 (Table 6, third column). The risk-factor LMH Community does have a positive sign (delta has a ”+” sign, Table 6, third column), i.e., there is a negative premium for high community relations firms, which would contradict the positive estimated effect in the cross-sectional approach. However, we interpret the results of this Charoenrook and Conrad (2005) test simply as evidence that the positive association found in Table 4 between community relations scores and stock returns was due to mispricing. Moreover, the benefits of good community relations might outweigh their costs, but stock prices might not properly discount this information since it is not publicly or widely available, which confirms the main working hypothesis and the mispricing scenario.

None of the other six ESG variables had any effect on stock returns, in either specification. By the no-effect scenario, this could be because the market values these ESG concerns efficiently, whether or not they carry any relevant information. However, a sign-shift in the effect of these ESG concerns on risk-adjusted returns during this period could also result in zero overall effect. Therefore, as robustness check of these findings, the effect of ESG performance on stock returns was estimated for two subperiods, 1992-2003 and 2003-2008. The 2003 breakpoint was chosen as there is reason to expect a change in ESG effect when more ESG information became available to investors. Table 5 reports estimation results for Equation (2) using the two plain ESG aggregation methods. The Best-in-Class method was dropped here as it has been shown that the inclusion of industry dummies yields similar estimates.

Human rights and product safety scores had consistent negative effects on stock returns in both periods, though statistically significant only in the later period. Conversely, community relations had a consistent positive effect but statistically significant only in the earlier period. The estimated economic magnitude of the marginal effect for a one-standard-deviation increase in the score (Relative or KO) is substantial: +0.59% or +0.71% per annum for community relations in 1992-2003, -0.92% or -2.18% per annum for human rights in 2003-2008, and -0.77% or -0.71% for product safety in 2003-2008. The aggregate ESG variable, while changing sign between the two periods was not statistically significant in either of them. More interestingly, the employee relations effect changed sign (and statistically significant) from positive during 1992-2003 to negative during 2003-2008. The economic magnitude of its effect is also substantial, shifting from 0.58 - 0.69% per annum in the earlier period to -0.56% or -0.64% per annum later (also for a one-standard-deviation increase in the score).

Next, Equation (3) was Garch-in-Mean estimated for the factor-mimicking portfolios for each of the ESG scores found to have an impact on stock returns, i.e., community relations and employee relations for 1992-2003, as well as human rights, product safety, and employee relations for 2003-2008 (Table 6). For simplicity, the Relative and Relative Best-in-Class aggregation methods for the ESG scores are used due to the strong similarity noted between theirs and their KO counterparts' behavior. The Best-in-Class method yields scores for which it is feasible to build the mutually exclusive mimicking portfolios at 30th- and 70th-percentile cutoff points.

However, with the Relative scores the 20th and 80th cutoff points had to be used for community relations and employee relations scores; 5th and 95th percentiles for human rights, and 15th and 85th percentiles for product safety.

The estimation results are presented in Table 6. Of all the candidate risk-factor-mimicking portfolios, only employee relations during 2003-2008 met the necessary condition of a positive relation between its conditional mean and conditional variance using either Relative or Best-in-Class scores (Table 6, second column). The positive sign of the relation indicates a positive premium for firms with low employee scores, i.e., a non-sustainability premium with respect to employee relations risk-factor, which confirms the second working hypothesis and the risk-factor scenario. Low employee relations scores are associated with high non-sustainability risk and therefore high expected returns. However, the intercept in the conditional mean equation is statistically different from zero, which suggests there are some idiosyncratic also driving the returns of the employee-relations portfolio.

A similar message emerges from Garch-in-Mean estimation on the employee-relations portfolio orthogonal to the Market, Small-Minus-Big, and High-Minus-Low portfolios (Table 7). The p-value of δ estimate and its sign indicate a strong positive association between its conditional mean and conditional variance in each of the specifications. However the intercept in the conditional mean equation still carries some extra information besides the conditional variance, which is weaker when using the Best-in-Class method. This could be an indication that some of the information which the intercept carries could be related to industry differences.

While there are reasons to further investigate the possibility that the negative effect of employee relations on returns during the later period might be due to compensation for non-sustainability risk, its positive effect in the earlier period is clearly due to mispricing. For the earlier period, the necessary condition is not met, as the p-value of the δ coefficient is higher than .10 (Table 6, first column). Thus the employee relations effect changed from positive, consistent with mispricing, to negative, possibly representing compensation for risk. This indicates that, during the earlier period, employee-relations benefits outweighted their costs.

In order for employee relations to have behaved as a risk factor in the later period, the sufficient condition of a Sharpe ratio estimate of 0.17 per month must also be met. Our Sharpe ratio estimates for this factor vary depending on the method used to compute the score, but are all substantially lower than the anticipated value, i.e., six or even ten times lower than the benchmark (Table 7, bottom line), meaning a 0.026 (Relative BC) or 0.017 (Relative) risk premium per month. In conclusion, as only the necessary condition for employee relations being a risk factor was met, there is some evidence that the higher returns earned by the low-employee-relations firms was compensation for risk.

The other factors, i.e., community relations in 1992-2003, and human rights and product safety in 2003-2008, did not satisfy the necessary condition to be risk-factors, i.e., the p-value of their δ estimates was higher than a conventional 0.10 (Table 6, first and second columns), which makes their effects on stock returns mispricing. Thus, the benefits of good community relations outweighted their costs, but the costs of good human rights and product safety outweighted their benefits.

Finally, the corporate governance, diversity, and environment scores had no statistically significant effects on risk-adjusted returns, neither during the overall period nor during the two subperiods. This seems to indicate either that information on these dimensions was efficiently reflected in stock returns or that it was not relevant at all.

Thus, we have found only weak evidence that employee relations could function as a risk factor (in the second period), but strong evidence that the other indicators' effects on stock returns were due to mispricing.

6. Summary and Conclusions

During the period July 1992- June 2008, only community relations had a positive effect on stock returns, due to mispricing. The benefits of having good community relations outweighted their costs, but this was not properly incorporated into stock prices.

Evidence of mispricing is also provided by the positive estimated effect of employee relations on stock returns during July 1992-June 2003, and the negative estimated effects of human rights and product safety on stock returns during July 2003-June 2008. The negative estimated effects for human rights and product safety indicate that their benefits were actually lower than their costs, but again this was not properly incorporated into stock prices.

There is weak evidence, however, that the change in the effect of employee relations from positive during June 1992-July 2003 to negative during July 2003-June 2008 could be due to compensation for risk. That is, firms with low scores in employee relations had higher expected stock returns than firms with high employee relations scores. This can be explained that firms with low scores in employee relations carry either a non-sustainability risk premium or a ”neglect” premium (by the discriminatory-tastes argument). This change in effect could be a consequence of better public information with respect to employee relations. However, during July 1992-June 2003, the market does not fully incorporate the net benefits of high employee relations.

Several problems with similar studies have been avoided here. First, in order to capture the actual relationship between ESG and stock returns, which might otherwise be dictated by the unobserved industry classification that could drive both ESG and stock returns, industry-specific effects were controlled for. The primary interest was thus in how ESG relates to stock returns within industries, not in the spurious relationship of returns and ESG across industries.

Second, improving over two frequently used alternative methods, a new aggregation method for measuring ESG was carefully designed and applied.

Third, because many empirical studies show that not all ESG dimensions are equally relevant for stock returns, or, more importantly, that there may be confounding effects among them, the effects of seven ESG dimensions – community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety – were analyzed separately.

These findings may be important for both investors and corporate strategists. Investors may be interested in new evidence that ESG performance is value relevant or that certain non-sustainability risks might exist, while firms might find that they can reduce their cost of capital by promoting certain ESG concerns.

APPENDIX A. ”GROUPING TECHNIQUE”

A problem when using the Fama-MacBeth two-pass procedure is estimation bias due to measurement error (sampling variance) in $\widehat{\beta}_j$, the errors-in-variables (EIV) problem (Black et al., 1972). The estimated $\widehat{\beta}_j$ in Equation 1 equals the true (unobservable) beta β_j plus a measurement error (or sample variance) v_j, assumed to be i.i.d. with zero mean and constant variance σ_v^2.

Therefore the residuals θ_j in Equation 1 also contain v_{jt}. OLS estimation leads to an inconsistent estimate of $\widehat{\gamma_1}$ due to the correlation between $\widehat{\beta}_j$ and the residuals θ_j. Even for large samples, as long as the variance of the measurement error σ_v^2 is positive, the estimated coefficient $\widehat{\gamma}_1$ will not converge to the true parameter γ_1 and will be biased towards zero, so-called attenuation bias (Greene, 2003, :85).

A solution developed by Black et al. (1972) and used by Fama and French (1992), the ”grouping technique”, provides N-consistent estimates of $\widehat{\gamma_1}$. It aims to reduce the variance in the residuals ϵ_j such that, when the sample size goes to infinity, this variance goes to zero.

Securities are first ordered in portfolios based on their individual beta estimates from the time series regressions

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \epsilon_{jt} \tag{7}$$

where R_{jt} is the excess return on asset j, month t; R_{mt} is the excess return on a market index, month t; $\alpha_j = E(R_j) - \beta_j E(R_m)$; $\beta_j = cov(R_{jt}, R_{mt})/var(R_{mt})$; and ϵ_j is independent of R_m and has a normal distribution with zero mean and constant variance.[7]

Portfolio excess return series are then constructed by taking the average of the excess returns in each portfolio. The estimation of Equation (7) - but now at portfolio level - yields *portfolio betas* instead of individual betas. In this case, as the individual error terms ϵ_j are by standard assumption i.i.d. in time and across assets, their averaging within portfolios results in a variance that is proportionally reduced by the number of securities in each portfolio. Thus, when the number of securities goes to infinity and the number of portfolios is kept fixed, the variance of the error terms converges to zero. The estimated coefficient $\widehat{\gamma_1}$ will then be N-consistent. Fama and MacBeth (1973), Fama and French (1992) and Amihud et al. (1992) acknowledge the benefits of this technique in providing more precise β_j estimates and used it in their empirical tests.

Due to their high correlation with the true but unobservable betas, individual beta estimates are used as ranking criteria. Firm-betas have been shown to be highly correlated with size as well (Fama and French, 1992), and therefore both size and individual beta estimates were used as ranking criteria.

For each year, all stocks were ranked in portfolios by firm size and individual β estimates. The individual β values were then estimated with the time series regression Equation (7) on 2 to 5 years of monthly observations prior to June of year t. By ranking the securities in 10 portfolios

[7]Because the market return includes asset i, the disturbances cannot be independent of R_m. However, Jensen (1969) shows that the resulting bias is extremely small (Friend and Blume, 1970).

based on market value as of December 31 of year $t-1$, and then separating each size-decile into five individual β breakpoints, 50 equally-weighted portfolios were formed as of June 30 of year t, and their returns were then computed for the next 12 months, through June of year $t+1$. Portfolio βs were estimated for each year based on the full remaining period and then assigned to the firms in that portfolio. Fama and French (1992) argue that "the precision of the full-period post-ranking portfolio βs, relative to the imprecise β estimates that would be obtained for individual stocks, more than makes up for the fact that the true βs are not the same for all stocks in a portfolio." At the same time, they note that variation in a firm's β is ensured by the fact that, due to changes in size or individual β estimate, a firm can change portfolios from year to year.

APPENDIX B. KLD'S ESG STRENGTH AND WEAKNESS INDICATORS

COMMUNITY STRENGTHS. Charitable Giving. The company has consistently given over 1.5% of trailing three-year net earnings before taxes (NEBT) to charity, or has otherwise been notably generous in its giving. Innovative Giving. The company has a notably innovative giving program that supports nonprofit organizations, particularly those promoting self-sufficiency among the economically disadvantaged. Companies that permit nontraditional federated charitable giving drives in the workplace are often noted in this section as well. Non-US Charitable Giving. The company has made a substantial effort to make charitable contributions abroad, as well as in the U.S. To qualify, a company must make at least 20% of its giving, or have taken notably innovative initiatives in its giving program, outside the U.S. 4. Support for Housing. The company is a prominent participant in public/private partnerships that support housing initiatives for the economically disadvantaged, e.g., the National Equity Fund or the Enterprise Foundation. Support for Education. The company has either been notably innovative in its support for primary or secondary school education, particularly for those programs that benefit the economically disadvantaged, or the company has prominently supported job-training programs for youth. Other Strength. The company has either an exceptionally strong volunteer program, in-kind giving program, or engages in other notably positive community activities. WEAKNESSES. Investment Controversies. The company is a financial institution whose lending or investment practices have led to controversies, particularly ones related to the Community Reinvestment Act. Negative Economic Impact. The companys actions have resulted in major controversies concerning its economic impact on the community. These controversies can include issues related to environmental contamination, water rights disputes, plant closings, put-or-pay contracts with trash incinerators, or other company actions that adversely affect the quality of life, tax base, or property values in the community. Other Weakness. The company is involved with a controversy that has mobilized community opposition, or is engaged in other noteworthy community controversies.

CORPORATE GOVERNANCE STRENGTHS. Limited Compensation. The company has recently awarded notably low levels of compensation to its top management or its board members. The limit for a rating is total compensation of less than $500,000 per year for a CEO or $30, 000 per year for outside directors. Ownership Strength. The company owns between 20% and 50% of another company KLD has cited as having an area of social strength, or is more than 20% owned by a firm that KLD has rated as having social strengths. When a company owns more than 50% of another firm, it has a controlling interest, and KLD treats the second firm as if it is a division of the first. Other Strength. The company has an innovative compensation plan for its board or executives, a unique and positive corporate culture, or some other initiative not covered by other KLD ratings. WEAKNESSES. High Compensation. The company has recently awarded notably high levels of compensation to its top management or its board members. The limit for a rating is total compensation of more than $10 million per year for a CEO or $100, 000 per year for outside directors. Tax Disputes. The company has recently been involved in major tax disputes involving more than $100 million with the Federal, state, or local authorities. Ownership Weakness. The company owns between 20% and 50% of a company KLD has cited as having an area of social weakness, or is more than 20% owned by a firm KLD has rated as having areas of weakness. When a company owns more than 50% of another firm, it has a controlling interest, and KLD treats the second firm as if it is a division of the first. Other Weakness. The company restated its earnings over an accounting controversy, has other accounting problems, or is involved with some other controversy not covered by other KLD ratings.

DIVERSITY STRENGTHS. CEO. The companys chief executive officer is a woman or a member of a minority group. Promotion. The company has made notable progress in the promotion of women and minorities, particularly to line positions with profit-and-loss responsibilities in the corporation. Board of Directors. Women, minorities, and/or the disabled hold four seats or more (with no double counting) on the board of directors, or one-third or more of the board seats if the board numbers less than 12. Work/Life Benefits. The company has outstanding employee benefits or other programs addressing work/life concerns, e.g., childcare, elder care, or flextime. Women & Minority Contracting. The company does at least 5% of its subcontracting, or otherwise has a demonstrably strong record on purchasing or contracting, with women- and/or minority-owned businesses. Employment of the Disabled. The company has implemented innovative hiring programs, other innovative human resource programs for the disabled, or otherwise has a superior reputation as an employer of the disabled. Gay & Lesbian Policies. The company has implemented notably progressive policies toward its gay and lesbian employees. In particular, it provides benefits to the domestic partners of its employees. Other Strength. The company has made a notable commitment to diversity that is not covered by other KLD ratings. WEAKNESSES. Controversies. The company has either paid substantial fines or civil penalties as a result of affirmative action controversies, or has otherwise been involved in major controversies related to affirmative action issues. Non-Representation. The company has no women on its board of directors or among its senior line managers. Other Weakness. The company is involved in diversity controversies not covered by other KLD ratings.

EMPLOYEE RELATIONS STRENGTHS. Cash Profit Sharing. The company has a cash profit sharing program through which it has recently made distributions to a majority of its workforce. Employee Involvement. The company strongly encourages worker involvement and/or ownership through stock options available to a majority of its employees, gain sharing, stock ownership, sharing of financial information, or participation in management decision-making. Health and Safety Strength. The company is noted by the US Occupational Health and Safety Administration for its safety programs. Retirement Benefits Strength. The company has a notably strong retirement benefits program. Union Relations. The company has a history of notably strong union relations. Other Strength. The company has strong employee relations initiatives not covered by other KLD ratings. WEAKNESSES. Union Relations. The company has a history of notably poor union relations. Health and Safety Weakness. The company recently has either paid substantial fines or civil penalties for willful violations of employee health and safety standards, or has been otherwise involved in major health and safety controversies. Workforce Reductions. The company has reduced its workforce by 15% in the most recent year or by 25% during the past two years, or it has announced plans for such reductions. Retirement Benefits Weakness. The company has either a substantially underfunded defined benefit pension plan, or an inadequate retirement benefits program. Other Weakness. The company is involved in an employee relations controversy that is not covered by other KLD ratings.

ENVIRONMENT STRENGTHS. Beneficial Products and Services. The company derives substantial revenues from innovative remediation products, environmental services, or products that promote the efficient use of energy [costa], or it has developed innovative products with environmental benefits. (The term environmental service does not include services with questionable environmental effects, such as landfills, incinerators, waste-to-energy plants, and deep injection wells.) Clean Energy. The company has taken significant measures to reduce its impact on climate change and air pollution through use of renewable energy and clean fuels or through energy efficiency. The company has demonstrated a commitment to promoting climate-friendly policies and practices outside its own operations. Communications. The company is a signatory to the CERES Principles, publishes a notably substantive environmental report, or has notably effective internal communications systems in place for environmental best practices. Pollution Prevention. The company has notably strong pollution prevention programs including both emissions reductions and toxic-use reduction programs. Recycling. The company either is a substantial user of recycled materials as raw materials in its manufacturing processes, or a major factor in the recycling industry. Other Strength. The company has demonstrated a superior commitment to management systems, voluntary programs, or other environmentally proactive activities. WEAKNESSES. Hazardous Waste. The companys liabilities for hazardous waste sites exceed $50 million, or the company has recently paid substantial fines or civil penalties for waste management violations. Regulatory Problems. The company has recently paid substantial fines or civil penalties for violations of air, water, or other environmental regulations, or it has a pattern of regulatory controversies under the Clean Air Act, Clean Water Act or other major environmental regulations. Ozone Depleting Chemicals. The company is among the top manufacturers of ozone depleting chemicals such as HCFCs, methyl chloroform, methylene chloride, or bromines. Substantial Emissions. The companys legal emissions of toxic chemicals (as defined by and reported to the EPA) from individual plants into the air and water are among the highest of the companies followed by KLD. Agricultural Chemicals. The company is a substantial producer of agricultural chemicals, i.e., pesticides or chemical fertilizers. Climate Change. The company derives substantial revenues from the sale of coal or oil and its derivative fuel products, or the company derives substantial revenues indirectly from the combustion of coal or oil and its derivative fuel products. Such companies include electric utilities, transportation companies with fleets of vehicles, auto and truck manufacturers, and other transportation equipment companies. Other Weakness. The company has been involved in an environmental controversy that is not covered by other KLD ratings.

HUMAN RIGHTS STRENGTHS. Indigenous Peoples Relations Strength. The company has established relations with indigenous peoples near its proposed or current operations (either in or outside the U.S.) that respect the sovereignty, land, culture, human rights, and intellectual property of the indigenous peoples. Labor Rights Strength. The company has outstanding transparency on overseas sourcing disclosure and monitoring, or has particularly good union relations outside the U.S. Other Strength. The company has undertaken exceptional human rights initiatives, including outstanding transparency or disclosure on human rights issues, or has otherwise shown industry leadership on human rights issues not covered by other KLD human rights ratings. WEAKNESSES. Burma Weakness. The company has operations or investment in, or sourcing from, Burma. Labor Rights Concern. The companys operations outside the U.S. have had major recent controversies related to employee relations and labor standards or its U.S. operations have had major recent controversies involving sweatshop conditions or child labor. Indigenous Peoples Relations Weakness. The company has been involved in serious controversies with indigenous peoples (either in or outside the U.S.) that indicate the company has not respected the sovereignty, land, culture, human rights, and intellectual property of indigenous peoples. Other Weakness. The companys operations outside the U.S. have been the subject of major recent human rights controversies not covered by other KLD ratings.

PRODUCT STRENGTHS. Quality. The company has a long-term, well-developed, company-wide quality program, or it has a quality program recognized as exceptional in U.S. industry. R&D/Innovation. The company is a leader in its industry for research and development (R&D), particularly by bringing notably innovative products to market. Benefits to Economically Disadvantaged. The company has as part of its basic mission the provision of products or services for the economically disadvantaged. Other Strength. The companys products have notable social benefits that are highly unusual or unique for its industry. WEAKNESSES. Product Safety. The company has recently paid substantial fines or civil penalties, or is involved in major recent controversies or regulatory actions, relating to the safety of its products and services. Marketing/ Contracting Controversy. The company has recently been involved in major marketing or contracting controversies, or has paid substantial fines or civil penalties relating to advertising practices, consumer fraud, or government contracting. Antitrust. The company has recently paid substantial fines or civil penalties for antitrust violations such as price fixing, collusion, or predatory pricing, or is involved in recent major controversies or regulatory actions relating to antitrust allegations. Other Weakness. The company has major controversies with its franchises, is an electric utility with nuclear safety problems, defective product issues, or is involved in other product related controversies not covered by other KLD ratings.

REFERENCES

Ahn DH, Conrad J, Dittmar RF. 2005. Basis assets. *Review of Financial Studies* **22**: 5133–5174.

Amihud Y, Christensen BJ, Mendelson H. 1992. Further evidence on the risk-return relationship. *Research Paper, Stanford University* **1248**: [Accessed 19.04.2010].

Arbel A, Carvell S, Strebel P. 1983. Giraffes, institutions and neglected firms. *Financial Analysts Journal* **39**: 2–8.

Barnea A, Rubin A. 2006. Corporate social responsibility as a conflict between shareholders. *EFA 2006 Zurich Meetings* **Available at** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 686606$: [Accessed 19.04.2010].

Bauer R, Koedijk K, Otten R. 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance* **29**: 1751–1767.

Becchetti L, Ciciretti R. 2006. Corporate social responsibility and stock market performance. *CEIS Working Paper* **79**: [Accessed 19.04.2010].

Black F, Jensen M, Scholes M. 1972. The capital asset pricing model: Some empirical tests. In Jensen M (ed.) *Studies in the Theory of Capital Markets*. Praeger Publishers, 72–194.

Charoenrook A, Conrad J. 2005. Identifying risk-based factors. *AFA 2006 Boston Meetings Paper* **Available at** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 757210$: [Accessed 19.04.2010].

Chatterji A, Brooks C, Pavelin S. 2008. How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy* **18**: 125–169.

Clark G, Knight E. 2009. Institutional investors, the political economy of corporate disclosure, and the market for corporate environmental and social responsibility: Implications from the uk companies act. *U. of Pennsylvania Journal of Business Law* **11**: 259–296.

Cochrane J. 2005. *Asset Pricing*. Princeton University Press.

Derwall J, Gunster N, Bauer R, Koedijk K. 2005. The ecoefficiency premium puzzle. *Financial Analysts Journal* **61**: 61–63.

Derwall J, Verwijmeren P. 2007. Csr and the cost of equity capital. In Derwall J (ed.) *The Economic Virtues of SRI and CSR*. Haveka, 193–213.

Dhaliwal D, Li OZ, Tsang A, Yang YG. 2009. Voluntary non-financial disclosure and the cost of equity capital: the case of corporate social responsibility reporting. *Working Paper* **Available at:** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 1343453$: [Accessed 19.04.2010].

Dowell G, Hart S, Yeung B. 2000. Do corporate global environmental standards create or destroy market value? *Management Science* **46**: 1059–1074.

Dufresne F, Savaria P. 2004. Corporate social responsibility and financial risk. *Journal of Investing* **13**: 57–66.

Edmans A. 2008. Does the stock market fully value intangibles? employee satisfaction and equity prices. *Working paper Series at the R.L. White Center for Financial Research* **16-07**:

[Accessed 19.04.2010].

Epstein M, Freedman M. 1994. Social disclosure and the individual investor. *Accounting, Auditing and Accountability Journal* **7**: 94–109.

Fama E, French K. 1992. The cross-section of expected returns. *Journal of Finance* **47**: 426–465.

Fama E, French K. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* **33**: 3–56.

Fama E, MacBeth J. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* **81**: 607–636.

Friend I, Blume M. 1970. Measurement of portfolio performance under uncertainty. *American Economic Review* **60**: 561–575.

Galema R, Plantinga A, Scholens B. 2008. The stocks at stake: return and risk in socially responsible investment. *Journal of Banking and Finance* **32**: 2646–2654.

Greene WH. 2003. *Econometric Analysis*. NJ Prentice Hall.

Haigh M, Hazelton J. 2004. Financial markets: A tool for social responsibility? *Journal of Business Ethics* **52**: 59–71.

Heinkel, Kraus A, Zechner J. 2001. The effect of green investment on corporate behavior. *Journal of Financial and Qantitative Analysis* **36**: 431–449.

Hong H, Kacperczyk M. 2009. The price of sin: the effects of social norms on markets. *Journal of Financial Economics* **93**: 15–36.

Jegadeesh N, Titman S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* **48**: 65–91.

Jensen M. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* **42**: 167–247.

Kempf A, Osthoff P. 2007. The effect of socially responsible investing on portfolio performance. *European Financial Management* **13**: 908–922.

Kinder P. 2005. Socially responsible investing: an evolving concept in a changing world **Available at** $http://www.kld.com/resources/papers/SRIevolving050901.pdf$: [Accessed 19.04.2010].

KPMG. 2008. International survey of corporate responsibility reporting **Available at** $http://us.kpmg.com/RutUS\_prod/Documents/8/Corporate\_Sustainability\_Report\_US\_Final.pdf$: [Accessed 19.04.2010].

MacKinley A. 1995. Multifactor models do not explain deviations from the capm. *Journal of Financial Economics* **38**: 3–28.

Merton R. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* **42**: 483–510.

Patten D. 1990. The market reaction to social responsibility disclosures: The case of the sullivan principles signing. *Accounting Organisations and Society* **15**: 515–87.

Renneboog L, Horst JT, Zhang C. 2008. Socially responsible investments: institutional aspects, performance, and investor behavior. *Journal of Banking and Finance* **32**: 1723–1742.

SIF. 2007. 2007 report on socially responsible investing trends in the united states, social investment forum - excutive summary **Available at** $http://www.socialinvest.org/pdf/SRI\_Trends\_ExecSummary\_2007.pdf$: [Accessed 19.04.2010].

Statman M. 2006. Socially responsible indexes. *Journal of Portfolio Management* **32**: 100–109.

Statman M, Glushkov D. 2009. The wages of social responsibility. *Financial Analysts Journal* **65**: 33–46.

Wall L. 1995. Some lessons from basic finance for effective socially responsible investing. *Economic Review - Federal Reserve Bank of Atlanta* **80**: 1–12.

TABLE 1. Mean, standard deviation (SD), and 25th-, 50th-, and 75th-percentile values for community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety scores and for an aggregate ESG score, calculated with the Relative and KO Aggregation as well as Best-in-Class methods, for the unbalanced panel of 10,174 firm-year observations during 1991-2006.

Panel 1: Relative Aggregation method (between -1 and 1)

	Community Rel.	Corp. Gov.	Diver-sity	Empl. Rel.	Env.	Human Rights	Prod. Safety	Aggreg. ESG
Mean	0.05	-0.10	0.04	0.02	-0.02	-0.03	-0.05	-0.02
SD	0.15	0.20	0.21	0.18	0.15	0.13	0.20	0.08
p25	0.00	-0.33	0.00	-0.03	0.00	0.00	0.00	-0.06
p50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
p75	0.14	0.00	0.12	0.17	0.00	0.00	0.00	0.04

Panel 2: KO Aggregation method (between 0 and 1)

Mean	0.44	0.41	0.33	0.49	0.51	0.70	0.48	0.48
SD	0.09	0.11	0.12	0.10	0.08	0.17	0.10	0.05
p25	0.40	0.33	0.27	0.40	0.50	0.57	0.50	0.45
p50	0.40	0.43	0.27	0.50	0.50	0.71	0.50	0.48
p75	0.50	0.50	0.36	0.54	0.54	0.80	0.50	0.51

Panel 3: Relative Best-in-Class method

Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD	0.14	0.18	0.20	0.18	0.14	0.12	0.19	0.08
p25	-0.06	-0.11	-0.08	-0.08	-0.02	0.00	-0.05	-0.04
p50	-0.03	0.02	0.00	0.00	0.00	0.01	0.02	0.00
p75	0.06	0.10	0.12	0.12	0.04	0.04	0.08	0.05

Panel 4: KO Best-in-Class method

Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD	0.08	0.09	0.11	0.09	0.07	0.10	0.10	0.04
p25	-0.04	-0.05	-0.07	-0.04	0.00	0.00	-0.02	-0.02
p50	-0.02	0.01	-0.01	-0.01	0.00	0.01	0.01	0.00
p75	0.03	0.04	0.06	0.06	0.02	0.03	0.04	0.03

TABLE 2. Mean, standard deviation (SD), 25th-, 50th- and 75th-percentile values for monthly Excess Returns; Estimated Beta; Book-to-Market Ratio; log of Market Equity; and Momentum factor; for the unbalanced panel of 9,202 firm-year observations during 1991-2006.

	Excess Returns (%)	Estimated Beta	Book-to-Market Ratio (%)	Market Equity (ln)	Momentum (%)
Mean	0.71	0.91	0.45	22.21	0.85
SD	9.71	0.36	0.34	1.57	3.05
p25	-4.39	0.61	0.24	21.21	-0.74
p50	0.65	0.87	0.38	22.26	0.85
p75	5.73	1.08	0.57	23.24	2.42

TABLE 3. Fama-MacBeth estimation of Model (1) over July 1992-June 2008, where the ESG variables are computed using Relative, KO, Relative Best-in-Class, and KO Best-in-Class methods. The dependent variable, excess returns, is in percentage points (%).

Without industry dummies	Relative		KO		Relative BC		KO BC	
Estimated Beta	0.08 (0.41)	0.07 (0.42)	0.08 (0.41)	0.07 (0.42)	0.07 (0.42)	0.06 (0.43)	0.07 (0.42)	0.05 (0.41)
Book-to-Market Ratio	0.42** (0.02)	0.38** (0.02)	0.43** (0.01)	0.39** (0.02)	0.44** (0.01)	0.41** (0.02)	0.44** (0.01)	0.41** (0.02)
Size	-0.01 (0.45)	-0.02 (0.34)	0.00 (0.47)	-0.02 (0.37)	0.00 (0.48)	-0.01 (0.40)	0.00 (0.48)	-0.01 (0.42)
Momentum	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)
Aggregate ESG	0.03 (0.47)		0.03 (0.47)		0.22 (0.32)		0.46 (0.30)	
Community Relations		0.12 (0.29)		0.36 (0.17)		0.24* (0.10)		0.43 (0.11)
Corporate Governance		-0.05 (0.40)		-0.26 (0.26)		-0.01 (0.48)		-0.13 (0.36)
Diversity		-0.06 (0.36)		-0.25 (0.24)		-0.10 (0.26)		-0.36 (0.13)
Employee Relations		0.14 (0.19)		0.36 (0.12)		0.13 (0.20)		0.31 (0.15)
Environment		-0.10 (0.35)		-0.23 (0.34)		-0.12 (0.29)		-0.24 (0.30)
Human Rights		-0.46 (0.08)		-0.70 (0.11)		-0.24 (0.21)		-0.46 (0.17)
Product Safety		-0.09 (0.33)		-0.21 (0.30)		-0.06 (0.37)		-0.11 (0.36)
Intercept	0.53 (0.33)	0.73 (0.23)	0.45 (0.37)	0.85 (0.17)	0.45 (0.35)	0.58 (0.28)	0.45 (0.35)	0.55 (0.29)
Mean Adj. R^2 Cross-section	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.07
Max. Adj. R^2 Cross-section	0.37	0.40	0.37	0.38	0.37	0.38	0.37	0.37
N	639	639	639	639	639	639	639	639
T	192	192	192	192	192	192	192	192

P-values in parentheses. *$p <= 0.10$, **$p <= 0.05$, ***$p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 4. Fama-MacBeth estimation of Model (2) over July 1992-June 2008, where the ESG variables are computed using Relative, KO, Relative Best-in-Class, and KO Best-in-Class methods. The dependent variable, excess returns, is in percentage points (%).

With industry dummies	Relative		KO		Relative BC		KO BC	
Estimated Beta	-0.01 (0.47)	-0.02 (0.47)	-0.02 (0.47)	-0.03 (0.46)	-0.01 (0.47)	-0.02 (0.47)	-0.02 (0.47)	-0.03 (0.46)
Book-to-Market Ratio	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)
Size	-0.02 (0.32)	-0.04 (0.19)	-0.02 (0.32)	-0.04 (0.20)	-0.02 (0.32)	-0.04 (0.19)	-0.02 (0.32)	-0.04 (0.20)
Momentum	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)
Aggregate ESG	0.13 (0.38)		0.30 (0.36)		0.13 (0.38)		0.30 (0.36)	
Community Relations		0.29* (0.05)		0.54* (0.05)		0.29* (0.05)		0.54* (0.05)
Corporate Governance		-0.02 (0.46)		-0.15 (0.34)		-0.02 (0.46)		-0.15 (0.34)
Diversity		-0.04 (0.39)		-0.26 (0.20)		-0.04 (0.39)		-0.26 (0.20)
Employee Relations		0.11 (0.24)		0.27 (0.18)		0.11 (0.24)		0.27 (0.18)
Environment		-0.21 (0.17)		-0.41 (0.17)		-0.21 (0.17)		-0.41 (0.17)
Human Rights		-0.26 (0.19)		-0.50 (0.15)		-0.26 (0.19)		-0.50 (0.15)
Product Safety		-0.12 (0.23)		-0.24 (0.23)		-0.12 (0.23)		-0.24 (0.23)
Intercept	1.49* (0.10)	1.65** (0.05)	1.34 (0.15)	1.69** (0.05)	1.49* (0.10)	1.67** (0.05)	1.49* (0.10)	1.64** (0.05)
Mean Adj. R^2 Cross-section	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Max. Adj. R^2 Cross-section	0.43	0.43	0.43	0.43	0.43	0.43	0.48	0.43
N	639	639	639	639	639	639	639	639
T	192	192	192	192	192	192	192	192

P-values in parentheses. $*p <= 0.10$, $**p <= 0.05$, $***p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 5. Fama-MacBeth estimation of Model (2) over July 1992-June 2003 and July 2003-June 2008, with industry dummies; the ESG variables are computed using Relative and KO methods. The dependent variable, excess returns, is in percentage points (%).

	1992-2003				2003-2008			
	Relative		KO		Relative		KO	
Estimated Beta	-0.06 (0.42)	-0.07 (0.41)	-0.07 (0.42)	-0.08 (0.40)	0.10 (0.37)	0.10 (0.37)	0.09 (0.37)	0.09 (0.38)
Book-to-Market Ratio	0.62*** (0.00)	0.59*** (0.00)	0.62*** (0.00)	0.59*** (0.00)	-0.27 (0.19)	-0.34 (0.13)	-0.26 (0.20)	-0.35 (0.12)
Size	0.00 (0.47)	-0.01 (0.44)	0.00 (0.48)	-0.01 (0.46)	-0.08 (0.12)	-0.11** (0.04)	-0.07 (0.13)	-0.11** (0.04)
Momentum	0.07** (0.01)	0.07** (0.01)	0.07** (0.01)	0.07** (0.01)	0.06* (0.08)	0.06* (0.08)	0.06* (0.08)	0.06* (0.08)
Aggregate ESG	0.52 (0.18)		1.13 (0.15)		-0.72 (0.13)		-1.53 (0.11)	
Community Relations		0.33* (0.08)		0.66* (0.06)		0.21 (0.23)		0.28 (0.30)
Corporate Governance		-0.06 (0.40)		-0.26 (0.30)		0.06 (0.39)		0.09 (0.43)
Diversity		-0.14 (0.25)		-0.44 (0.14)		0.16 (0.24)		0.15 (0.35)
Employee Relations		0.27* (0.09)		0.64** (0.05)		-0.26* (0.07)		-0.53* (0.07)
Environment		-0.14 (0.30)		-0.29 (0.31)		-0.36 (0.13)		-0.68 (0.14)
Human Rights		-0.11 (0.38)		-0.25 (0.34)		-0.59* (0.10)		-1.07* (0.10)
Product Safety		-0.03 (0.44)		-0.08 (0.42)		-0.32* (0.09)		-0.59* (0.10)
Intercept	0.37 (0.40)	0.63 (0.34)	-0.13 (0.47)	0.59 (0.37)	3.97** (0.02)	4.78*** (0.00)	4.59** (0.01)	6.10*** (0.00)
Mean Adj. R^2 Cross-section	0.13	0.14	0.13	0.14	0.13	0.13	0.13	0.13
Max. Adj. R^2 Cross-section	0.43	0.43	0.43	0.43	0.37	0.38	0.38	0.38
N	639	639	639	639	639	639	639	639
T	132	132	132	132	60	60	60	60

P-values in parentheses. $*p <= 0.10$, $**p <= 0.05$, $***p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 6. P-value and sign for the conditional variance (delta) and intercept (mu) parameter estimates in the conditional mean equation of the Garch-in-Mean model (Equation 3) for risk-factor mimicking portfolios over July 1992-June 2003, July 2003-June 2008 and July 1992-June 2008. The risk-factor mimicking portfolios are built as Low minus High ESG scores, computed using both Relative and Relative Best-in-Class methods. In the Relative Best-in-Class approach, 30th- and 70th-percentile breakpoints were used; in the Relative method, 20th- and 80th-percentiles for community relations and employee relations; 5th- and 95th-percentiles for human rights; and 15th- and 85th-percentiles for product safety.

		1992-2003				2003-2008				1992-2008			
		Relative		Relative BC		Relative		Relative BC		Relative		Relative BC	
		p-val	sign	p-val	sign	p-val	sign	p-val	sign	p-val	sign	p-val	sign
Community Relations	delta	0.38	+	0.35	+	0.33	+	0.47	-	0.32	+	0.40	+
	mu	0.17	-	0.28	-	0.43	-	0.48	+	0.22	-	0.21	-
Employee Relations	delta	0.22	+	0.32	+	0.00***	+	0.05**	+	0.29	+	0.28	+
	mu	0.17	-	0.27	-	0.00***	-	0.09*	-	0.27	-	0.34	-
Human Rights	delta	0.34	+	0.45	+	0.15	-	0.26	-	0.15	+	0.38	+
	mu	0.24	-	0.11	-	0.11	+	0.40	+	0.17	-	0.10*	-
Product Safety	delta	0.39	+	0.30	+	0.13	-	0.18	-	0.48	+	0.31	+
	mu	0.44	-	0.42	+	0.14	+	0.13	+	0.48	-	0.37	+
T		132		132		60		60		192		192	

$*p <= 0.10$, $**p <= 0.05$, $***p <= 0.01$. T=number of monthly observations.

TABLE 7. Parameter estimates in the conditional mean and conditional variance equation of the Garch-in-Mean model in Equation (3) for the employee-relations risk-factor mimicking portfolio, both plain and orthogonal to the Market, Small-Minus-Big size (SMB) and High-Minus-Low (HML) risk-factors, over July 2003-June 2008. The monthly returns of the employee relations portfolio are in percentage points. The employee relations score was computed using the Relative and Relative Best-in-Class (Relative BC) methods. In the Relative Best-in-Class method, 30th- and 70th-percentile breakpoints were used; in the Relative method, 20th and 80th percentiles.

	Relative				Relative BC			
		Orthogonal to				Orthogonal to		
	Plain	Market	SMB	HML	Plain	Market	SMB	HML
Mean Equation								
intercept	-1.31**	-1.66*	-1.50**	-1.18**	-2.58*	-2.71**	-2.13*	-2.70*
	(0.00)	(0.10)	(0.00)	(0.02)	(0.10)	(0.03)	(0.10)	(0.07)
delta	1.52***	1.79*	1.67***	1.16**	2.29*	2.24***	2.02**	2.33**
	(0.00)	(0.05)	(0.00)	(0.01)	(0.05)	(0.01)	(0.04)	(0.03)
Conditional Variance Equation								
intercept	0.35***	0.33**	0.43***	0.00	0.18	0.00	0.13	0.13
	(0.00)	(0.04)	(0.00)	(0.50)	(0.16)	(0.50)	(0.39)	(0.26)
ARCH1	0.49***	0.63***	0.44***	0.90***	0.83***	0.98***	0.86**	0.87***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)
GARCH1	0.25***	0.09*	0.24***	0.14***	0.03*	0.02*	0.04	0.03*
	(0.00)	(0.09)	(0.00)	(0.00)	(0.10)	(0.08)	(0.13)	(0.07)
T	60	60	60	60	60	60	60	60
Sharpe Ratio (%)	1.66	1.95	1.85	1.30	2.59	2.64	2.23	2.67

P-values in parentheses. *$p <= 0.10$, **$p <= 0.05$, ***$p <= 0.01$. T=number of monthly observations.