Review of Financial Economics 26 (2015) 25–35

Contents lists available at ScienceDirect

Review of Financial Economics

journal homepage: www.elsevier.com/locate/rfe

The wages of social responsibility — where are they? A critical review of ESG investing

Gerhard Halbritter *, Gregor Dorfleitner [1]

University of Regensburg, Department of Finance, Universitätsstraße 31, 93053 Regensburg, Germany

ARTICLE INFO

Article history:
Received 22 December 2014
Received in revised form 18 February 2015
Accepted 24 March 2015
Available online 2 April 2015

JEL classification:
G10
G11
G12

Keywords:
ESG
Environmental social governance
Socially responsible investing
Corporate social performance
Financial performance

ABSTRACT

This paper contributes both to investigating the link between the corporate social and financial performance based on environmental, social and corporate governance (ESG) ratings and to reviewing the existing empirical evidence pertaining to this relationship. The sample used includes ESG data of ASSET4, Bloomberg and KLD for the U.S. market from 1991 to 2012. The econometrical framework applies an ESG portfolio approach using the Carhart (1997) four-factor model as well as cross-sectional Fama and MacBeth (1973) regressions. Previous empirical research indicates a relationship between ESG ratings and returns. As against this, the ESG portfolios do not state a significant return difference between companies with high and low ESG ratings. Although the Fama and MacBeth (1973) regressions reveal a significant influence of several ESG variables, investors are hardly able to exploit this relationship. The magnitude and direction of the impact are substantially dependent on the rating provider, the company sample and the particular subperiod. The results suggest that investors should no longer expect abnormal returns by trading a difference portfolio of high and low rated firms with regard to ESG aspects.

1. Introduction

Several studies examine the relationship between a company's social and financial performance based on environmental, social and corporate governance ratings, or short ESG ratings, mostly finding a positive connection. This paper contributes to this subject by revisiting the evidence from the perspective of a U.S. investor utilizing the data of three different ESG ratings. By analyzing the problem with a sample consisting of data up to 2012 and by applying an econometrically sound methodology, we show that the positive effects reported earlier are both dependent on the rating concept used and the time interval in which the observations were made.

Throughout the last decade, Socially Responsible Investments (SRIs) have experienced an impressive development. According to the U.S. Forum for Sustainable and Responsible Investments USSIF (2012), SRIs account for more than 20% of the global capital market. Assets exceeding $30 trillion are based on SRI principles. These figures reveal the outstanding importance for both investors and researchers. Apart from the corporate financial performance (CFP), private and institutional investors are increasingly interested in the corporate social performance (CSP) of a particular company. Firms are being ever more encouraged to consider non-monetary goals.

Nevertheless, most investors consider social concerns just as a side condition while a maximization of the return is still the primary objective. In this context the question of whether there is a link between the financial and social performance of a company arises. Even investors without non-monetary interests could exploit a potential relationship to gain abnormal returns. A meta-analysis of Orlitzky, Schmidt, and Rynes (2003) finds that the CFP is positively correlated with the CSP while the dependence is bidirectional and simultaneous. Wallis and Klein (in press) also suggest that there is a certain amount of evidence for an outperformance of socially responsible over conventional investments.

The oldest line of SRI research is concerned with a comparison of the performance of conventional and SRI funds. Most studies in this field, such as Hamilton, Jo, and Statman (1993), Statman (2000), Bauer, Koedijk, and Otten (2005), Bello (2005), Kreander, Gray, Power, and Sinclair (2005), and Utz and Wimmer (2014), do not indicate significant performance differences. A number of papers, such as Sauer (1997), Statman (2000), Schröder (2004), Statman (2006), Schröder (2007), and Lee and Faff (2009), do not provide evidence of an out- or underperformance of SRI indices compared to conventional indices.

* Corresponding author. Tel.: +49 941 943 2684.
E-mail addresses: Gerhard.Halbritter@ur.de (G. Halbritter), Gregor.Dorfleitner@ur.de (G. Dorfleitner).
[1] Tel.: +49 941 943 2684.

http://dx.doi.org/10.1016/j.rfe.2015.03.004
1058-3300/

A recent analysis of Belghitar, Clark, and Deshmukh (2014) finds that there is no difference regarding the expected returns and their variance. However, SR investors pay a high price in terms of utility if higher moments are taken into account.

Utz and Wimmer (2014) state that SRI mutual funds do not, on average, hold socially responsible firms to a greater extent than conventional funds do. As a result, it becomes questionable whether packaged SRI products are at all suitable for investigating the link between the social and the financial performance. In this context, ESG ratings allow for a more appropriate approach, as they provide a direct measure of the CSP at company level. Specialized rating agencies define a certain set of criteria incorporating a variety of sustainability issues. Each corporation in the rating universe receives a specific rating score. In contrast to SRI funds and indices, ESG ratings result in large panel data sets which offer a considerably more precise understanding of how sustainability aspects influence a firm's return. The related literature discussed in Section 2 provides evidence that investors are able to attain abnormal returns by trading corporations with high ESG scores long and low ESG scores short.

This paper contributes to reviewing this existing empirical evidence based on ESG ratings pertaining to the link between the CSP and CFP of a company. Most studies focus on one special ESG rating database. However, it is possible that implications related to returns are dependent on the underlying rating approach. Dorfleitner, Halbritter, and Nguyen (2014) reveal significant differences in distribution, level and risk of various ESG rating concepts. Furthermore, most studies are based on very short time-series since most rating agencies did not commence their work before the beginning of the last decade.

This is the first paper to compare the impact of sustainability issues on a firm's return with respect to three ESG rating providers, namely ASSET4, Bloomberg and KLD. As a part of Thomson Reuters, ASSET4 is one of largest ESG rating agencies. The data set starts in 2002 and includes more than 1000 U.S. companies. The Bloomberg sample starts in 2005 and also contains over 1000 U.S. firms. Similar studies mostly apply the ESG scores according to Kinder, Lydenberg, and Domini Research & Analytics (KLD). With more than 4000 firms and a history starting in 1990, KLD offers the most comprehensive ESG database. The relatively large sample allows us both to gain a profound insight into the relationship of the financial and the social performance and to compare the three ESG approaches in terms of return predictability.

The empirical framework is guided by existing research approaches presented in Section 2 and by econometrical adequacy. We follow two different strategies — an ESG portfolio method in the spirit of Kempf and Osthoff (2007) and cross-sectional regressions similar to Galema, Plantinga, and Scholtens (2008). ESG portfolios provide a straightforward strategy for investors to exploit a potential relationship between ESG ratings and the financial performance while the cross-sectional regressions allow us to obtain a more profound understanding of how the ESG level affects the return in the cross-section.

The ESG portfolio method generally constructs a *high* and a *low* portfolio including ESG out- and underperformers, respectively. Additionally, a *high–low* difference portfolio buys the high portfolio and short sells the low one. We use the Carhart (1997) four-factor model for estimating abnormal returns. A best-in-class approach instead uses sector-specific ESG ratings for composing the portfolios. In order to exploit the full sample size and to avoid assumptions regarding the portfolio construction, we also consider monthly cross-sectional Fama and MacBeth (1973) regressions.

The remainder of this paper is organized as follows. Section 2 outlines the current state of research. Section 3 presents the data set. Section 4 develops the empirical framework for our analysis while Section 5 presents the results and implications. Finally, Section 6 concludes.

2. Related literature

In this section we provide an overview of the literature concerning the link between a firm's return and CSP level, measured in terms of ESG criteria. All papers investigating this issue for the U.S. market by using ESG portfolios are considered. Moreover, we review research applying cross-sectional regressions in order to identify the explanatory power of ESG variables without relying on portfolio assumptions.

Based on the ratings of Innovest, Derwall, Guenster, Bauer, and Koedijk (2005) investigate whether ecological responsibility has an impact on a company's return. The sample includes U.S. companies for the years 1997 to 2003. By applying a high–low strategy, the Carhart (1997) four-factor model reveals a significant outperformance of high-rated firms over low-rated ones.

Eccles, Ioannou, and Serafeim (2014) follow a combined approach to identify high and low sustainability firms from a sample of 180 U.S. companies. Beside the ESG ratings of ASSET4 and SAM, they also rely on personal research and interviews. Based on the general impression, half of the firms are categorized as being high or rather low. The Carhart (1997) four-factor model following a high–low strategy reveals annual abnormal returns of up to 4.8% for a sample period from 1993 to 2010.

Lee, Faff, and Rekker (2013) also investigate the performance of U.S. companies dependent on the ESG ratings of SAM. The Carhart (1997) four-factor model for 1998 to 2007 provides evidence in favor of a significant outperformance of high-rated companies as well as of high-rated sectors. As opposed to our work, the study only accounts for an overall ESG score and not for the particular ESG pillars.

Since KLD was one of the first ESG rating providers, their database is relatively large and starts early in 1990. They also use a very transparent scoring approach based on a number of strengths and concerns. For this reason, KLD data is used in numerous empirical papers. As one of the first studies based on KLD ratings, Kempf and Osthoff (2007) compare the performance of high- and low-rated companies from the S&P 500 and DS 400 for the years 1992 to 2004. The value-weighted portfolios are constructed by using a 10% cut-off.[2] The Carhart (1997) four-factor model reveals a significant performance difference as related to the high–low portfolio. Investors were able to realize an abnormal return of up to 8.7% per year.

Statman and Glushkov (2009) also compose high and low portfolios based on KLD rating data from 1992 to 2007. As opposed to Kempf and Osthoff, (2007), their portfolios are equally-weighted and the cut-off is one third. For most ESG categories, both the CAPM and the Carhart (1997) four-factor model indicate a significant positive abnormal return of a high–low strategy.

In contrast to previous studies, Galema et al. (2008) construct portfolios separately for each KLD strength and concern. All strength-minus-concern portfolios demonstrate a positive Carhart (1997) four-factor model alpha using data from 1992 to 2006. By estimating Fama and MacBeth (1973) regressions, they conclude that the employee relations indicator has a significant positive effect on the return.

Manescu (2011) examines the connection between ESG scores and returns in the cross-section based on KLD data from 1991 to 2006 including all S&P 500 and DS 400 companies. Fama and MacBeth (1973) regressions demonstrate that the community relations criteria have a significant positive influence on the return. The overall ESG rating does not explain returns.

In summary, the findings of the empirical research provide evidence in favor of a link between returns and the CSP level, measured in terms of the environmental, social and corporate governance dimensions. Based on ASSET4, Innovest, KLD and SAM ratings, Derwall et al. (2005), Eccles et al. (2014), Kempf and Osthoff (2007) and Statman and Glushkov (2009) find a significant positive impact of the ESG

[2] In the context of ESG portfolios the cut-off generally describes the top and bottom quantile of companies considered as high and low, respectively.

score on the return. This result suggests that investors can gain abnormal returns by trading ESG difference portfolios. While these studies only apply the ESG portfolio approach, Galema et al. (2008) and Manescu (2011) account for a dependence in the cross-section. Both studies show a significant positive relationship between the ESG and return variables for at least a few indicators.

Although there is evidence of the fact that the financial performance of a company is dependent on its ESG score, this issue must be critically reviewed. Most studies are based on one specific ESG data set. According to Dorfleitner, Halbritter, and Nguyen (2014), the ESG ratings of ASSET4, Bloomberg and KLD are significantly different in terms of both distribution and risk. This aspect may also affect a potential correlation with financial items. For this reason it is crucial to account for different ESG rating providers. As a consequence, our study uses multiple ESG rating concepts. We consider both the overall ESG level and the pillars in terms of the environmental, social and governance performance. Moreover, due to the fact that our database carries on until 2012, we are able to include recent developments in our analysis. Compared to the studies of Galema et al. (2008) and Manescu (2011), we account for autocorrelations of the time-variant Fama and MacBeth (1973) coefficients as proposed by Fama and French (2002) and Petersen (2009).

3. Data

The data set covers the full ASSET4 rating universe for U.S. companies from 2002 to 2011. The rating approach is based on more than 850 indicators with regard to ESG aspects. Within various steps these indicators are condensed to four pillars relating to the environmental, social, governance and economic performance. As a part of the scoring process, all companies are benchmarked against the complete firm universe. ASSET4 also provides an overall ESG score composed of the equally weighted pillars.

From Bloomberg we only acquired ESG scores for firms which are also included in the ASSET4 database. The sample covers the years from 2005 to 2011. The rating approach of Bloomberg incorporates over 100 data points. Similar to ASSET4, these indicators are aggregated to a total ESG score and the three sub-categories environment, social and governance.

We fully include the database of KLD except for privately held companies. As KLD is one of the first ESG rating providers, the sample covers the years from 1990 to 2011. Compared to ASSET4 and Bloomberg, KLD does not offer numerical ESG scores but rather binary indicators for numerous strengths and concerns. The total number of indicators is variable over time. In order to make the KLD ratings compatible with the scores of ASSET4 and Bloomberg, the strengths and concerns need to be transformed to numerical values. On this account we follow Kempf and Osthoff (2007) and revert all concerns back into strengths by taking the opposite binary value. For each pillar environment, social and governance we sum up the particular indicators and normalize them between 0 and 100. Analogously, the total ESG score is calculated by using all indicators. Additionally, KLD provides data whether companies are involved in controversial business sectors, namely alcohol, firearms, gambling, military, nuclear power, and tobacco. We therefore calculate an alternative total ESG* score including these indicators in the sense of concerns.

As a result, we have a comparable data set of three ESG rating providers including variables for the total ESG score (ESG) as well as for the individual score of the pillars environment (ENV), social (SOC) and governance (GOV). In addition, ASSET4 reports upon a fourth subscore related to a firm's economic sustainability (ECN). For KLD we also account for a total ESG score including negative screening criteria (ESG*).

Even though all three rating agencies offer a measure for a firm's CSP, their data sets feature notable differences. Dorfleitner, Halbritter, and Nguyen (2014) find significant variations in distribution and risk characteristics. Table 1 shows the descriptive statistics of the ESG data.[3] Concerning ASSET4, the mean ESG scores of the entire rating universe are supposed to be in a range of 50 due to their scoring approach. Considering only U.S. companies, the mean ESG and ECN values are consistent with this assumption. As opposed

Table 1
Descriptive statistics: ESG ratings.

			Mean	SD	Min	Max	Observations	
ASSET4	ESG	Overall	51.87	28.19	3.02	98.57	N	88,600
		Between		23.62	4.23	97.44	n	1170
		Within		15.19	−8.84	114.00	$\overline{T}$	75.73
	ENV	Overall	40.05	30.61	8.90	97.16	N	88,600
		Between		25.16	9.10	96.59	n	1170
		Within		16.37	−30.07	108.64	$\overline{T}$	75.73
	SOC	Overall	44.30	28.41	3.68	98.88	N	88,600
		Between		23.80	4.47	95.61	n	1170
		Within		15.22	−20.10	104.87	$\overline{T}$	75.73
	GOV	Overall	73.20	16.94	1.50	97.94	N	88,600
		Between		13.59	7.50	95.28	n	1170
		Within		11.52	3.26	118.86	$\overline{T}$	75.73
	ECN	Overall	52.09	28.32	1.21	98.98	N	88,600
		Between		21.31	2.09	96.65	n	1170
		Within		19.37	−15.51	117.67	$\overline{T}$	75.73
Bloomberg	ESG	Overall	21.07	11.94	0.75	85.12	N	56,721
		Between		9.98	6.61	60.96	n	1073
		Within		5.04	−20.73	50.76	$\overline{T}$	52.86
	ENV	Overall	20.55	15.84	0.78	89.92	N	27,135
		Between		13.26	0.78	61.37	n	585
		Within		7.65	−22.34	53.42	$\overline{T}$	46.38
	SOC	Overall	15.42	15.12	3.13	83.33	N	53,606
		Between		12.60	3.13	65.61	n	986
		Within		6.77	−33.28	55.19	$\overline{T}$	54.37
	GOV	Overall	52.89	6.23	5.36	85.71	N	56,673
		Between		5.49	16.52	72.50	n	1073
		Within		3.17	18.61	70.53	$\overline{T}$	52.82
KLD	ESG	Overall	67.28	12.04	25.58	100.00	N	342,457
		Between		8.25	34.23	93.35	n	4209
		Within		9.72	20.40	95.19	$\overline{T}$	81.36
	ESG*	Overall	71.56	10.81	32.65	100.00	N	342,457
		Between		7.37	42.63	94.30	n	4209
		Within		8.62	30.54	95.77	$\overline{T}$	81.36
	ENV	Overall	64.09	13.66	9.09	100.00	N	342,457
		Between		9.25	18.18	88.89	n	4209
		Within		11.06	3.04	102.75	$\overline{T}$	81.36
	SOC	Overall	61.01	11.16	17.86	100.00	N	342,457
		Between		7.41	32.14	92.83	n	4209
		Within		8.15	17.07	92.40	$\overline{T}$	81.36
	GOV	Overall	62.82	17.40	0.00	100.00	N	342,457
		Between		10.67	33.33	100.00	n	4209
		Within		15.07	4.02	117.79	$\overline{T}$	81.36

This table presents the mean, standard deviation, minimum and maximum values of the ESG scores separated into providers and pillars. *Overall* denotes the full data set. *Between* indicates the cross-section while *within* describes the time-series dimension. *N* is the number of total observations for *n* companies over an average time-series of $\overline{T}$ months.

[3] The standard deviations are calculated as follows:

$$\text{Overall}: s_o = \sqrt{\frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t}-\overline{x}\right)^2}.$$

$$\text{Between}: s_b = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\overline{x}_i-\overline{x}\right)^2}.$$

$$\text{Within}: s_w = \sqrt{\frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{i,t}-\overline{x}_i\right)^2}.$$

Minimum and maximum values are based on the following variables:

$$\text{Overall}: x_{i,t}.$$
$$\text{Between}: \overline{x}_i.$$
$$\text{Within}: x_{i,t}-\overline{x}_i+\overline{x}.$$

to this, the GOV rating exhibits significantly higher values of approximately 75. This indicates that U.S. companies are above average regarding corporate governance practices. On the other hand, the environmental and social dimensions are slightly below average. The mean Bloomberg scores are located on a lower level while KLD's means are clearly above 60. The average scores are even higher if controversial business involvement data is incorporated. This suggests that the majority of the companies in the sample are related to indisputable sectors. The overall standard deviation of Bloomberg and KLD ranges between 10 and 20 while ASSET4 exhibits notably higher ones around 30. Regarding ASSET4 and Bloomberg, the variation between the firms is twice as big as the temporal variation within one company. The KLD ratings show lower volatilities which are fairly in line with each other.

The KLD data set features 342,457 total observations (N) based on 4209 U.S. companies (n) with an average time-series of 81 months ($\bar{T}$). Since ASSET4 and Bloomberg do not solely provide ratings for U.S. firms, they focus on larger corporations. ASSET4 offers scores for 1170 U.S. enterprises with a mean history of 76 months. The Bloomberg data consists of 1073 companies and an average time-series of 53 months. 996 firms are covered by all three agencies.

As this analysis takes the perspective of a U.S. investor, the risk-free interest rate is represented by the one-month U.S. Treasury bill and all data are denominated in U.S. Dollars. Total returns, market capitalization data and book-to-market ratios are retrieved from Thomson Reuters Datastream. Discarded or insolvent companies are included up to the last available financial or rating information. In the case of a merger, both companies are devolved upon the firm retaining its Datastream ticker. Therefore, our study is not subject to a survivorship bias.

In the context of the Carhart (1997) four-factor model, we also need risk premia relating to size, value and momentum. Research commonly uses risk factors available from Kenneth French's data library. However, Asness and Frazzini (2013) criticize the construction of the high-minus-low factor (HML). Forming the book-to-market value portfolios, Fama and French (1992) use data that is up to 18 months old. As a consequence, the value measure is not unraveled from the momentum anomaly determined on a monthly interval. Based on a more timely measurement, Asness and Frazzini (2013) account for a true value strategy, not a value-momentum strategy. For this reason we apply the risk premia provided on the website of Andrea Frazzini.[4] These risk premia are constructed in accordance with Fama and French (1992, 1993), Carhart (1997) and Asness and Frazzini (2013).

Within the framework of the panel regressions, sector data is required to account for intra-industry dependencies. For this purpose, the Standard Industrial Classification (SIC) is used. The data set includes 653 different SIC classes. For the best-in-class ratings the companies are only grouped into the ten major sectors of basic materials, consumer cyclicals, consumer non-cyclicals, energy, financials, healthcare, industrials, technology, telecommunication services and utilities. This ensures a sufficiently high number of firms in each class.

4. Methodology

4.1. ESG portfolios

As the empirical literature illustrates, constructing ESG portfolios is one of the most common approaches to investigating the relationship between the social and financial performance of companies. Based on ESG ratings, this method easily aggregates large panel data sets to a single time-series dimension. This allows the application of basic asset pricing models. Furthermore, it provides a straightforward trading strategy to investors. The empirical framework of this section largely complies with Kempf and Osthoff (2007) and Statman and Glushkov (2009).

In each year p from 1991 to 2012 we construct two market capitalization weighted portfolios for each ESG rating of the three data providers. For this purpose, we sort the companies according to the score available at p-1. The 20% best (worst) performing firms in a particular category are assigned to the *high* (*low*) portfolio. In order to compare the performance of both portfolios, we investigate a *high–low* strategy which contains the high portfolio in a long position while the low portfolio is held in a short position.

To evaluate the performance of these ESG portfolios, we apply the Carhart (1997) four-factor model. The model is estimated given the following equation:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i\left(r_{m,t} - r_{f,t}\right) + s_i SMB_{i,t} + h_i HML_{i,t} + w_i WML_{i,t} + u_{i,t}, \tag{1}$$

where $r_{i,t} - r_{f,t}$ is the excess return of the portfolio i over the risk-free interest rate in month t. It is explained by the excess return of the market $r_{m,t} - r_{f,t}$ as well as by the size, value and momentum factor denoted as $SMB_{i,t}$, $HML_{i,t}$, and $WML_{i,t}$, respectively. The coefficients α_i, β_i, s_i, h_i, and w_i are estimated performing a linear regression while $u_{i,t}$ represents the residual. Standard errors are calculated using the Newey and West (1987) procedure.[5]

For each ESG score type we consider both the high and low portfolio and the high–low strategy. In terms of the latter, the alpha expresses the return difference between ESG outperformers and underperformers. In a first step, we allow for the full existing data set. To control for firm-specific effects, we restrict the sample to companies and observations available from all three providers in a second step.

In order to gain further robustness, we also estimate the full sample model by applying various modifications. Accounting for sector-specific issues we also test a best-in-class version. The best-in-class score of a company is calculated as the difference of the individual and the average industry score. During the portfolio selection process firms are by implication ranked among their own peer group. As a consequence, companies can also be eligible for a high portfolio even if they belong to a sector which is difficult in terms of ESG requirements.

To investigate whether the results are dependent on the assumptions of the portfolio selection process, we also consider variations in the cut-off and weighting definition. In addition to the initial cut-off of 20%, we incorporate a 1, 5, 10, 25 and 50% barrier, each in a market capitalization and equally weighted implementation.

Furthermore, we examine whether the link between the financial and social performance remains constant over time. The sample is split into the three subperiods from 1991 to 2001, from 2002 to 2006, and from 2007 to 2012. The first 10 years of the full sample period are only applicable to the KLD data set. Since we are interested in a comparison of the three rating providers, we do not further split this period. In the second decade we also consider rating data of ASSET4 and Bloomberg. In this case, shorter periods allow us to track the return development based on the ESG ratings of all three providers.

The portfolio is constructed both in a market capitalization and equally weighted version. Given the core issue of this paper, all robustness checks focus on the alphas of the high–low strategy as a measure of the abnormal performance.

4.2. Cross-sectional regressions

In the previous section, the relationship of the CSR on the financial performance is examined aggregating the cross-sectional dimension to portfolios. In this section, we apply a panel-based strategy to analyze

[4] See: http://www.econ.yale.edu/~af227/data_library.htm.

[5] A Breusch and Pagan (1979) test performed on all models provides evidence of the fact that the residuals of the linear regressions are subject to heteroskedasticity. A Durbin and Watson (1971) test indicates autocorrelations for some of the models. As a robustness check, we also estimate the regressions using the conventional OLS estimator. The implications remain unchanged.

the direct impact of ESG variables on the return using Fama and MacBeth (1973) regressions. While the portfolio approach only includes companies with a very high and very low ESG score, this procedure accounts for the full cross-section without the necessity to make assumptions about the portfolio construction.

Fama and MacBeth (1973) regressions are designed to measure the influence of the systematic risk factor beta on the return of a company. Fama (1991) and Fama and French (1992) emphasize that both the market capitalization and the book-to-market ratio were also important for explaining stock returns in the cross-section. In this context ESG scores could also contribute.

We estimate two model specifications. While model (I) uses the overall ESG score as an explanatory variable, model (II) focuses on the particular ESG pillars. The betas are estimated following the two-step sorting procedure of Black, Jensen, and Scholes (1972). A vector of control variables is also included. Following Galema et al. (2008) and Hong and Kacperczyk (2009), we use the natural logarithm of the market capitalization and the book-to-market ratio as well as the average return over the last 12 months.

For each month t we estimate the following cross-sectional regression:

$$r_{i,t}-r_{f,t} = \gamma_{0,t} + \gamma_{1,t}\hat{\beta}_{i,t} + \mathbf{X}_{i,t-1}\ \boldsymbol{\gamma}_{X,t} + \mathbf{ESG}_{i,t-1}\ \boldsymbol{\gamma}_{ESG,t} + u_{i,t} \quad \forall \quad t = 1,2,\ldots,T, \tag{2}$$

where $r_{i,t} - r_{f,t}$ is the return of company i over the risk-free interest rate. $\hat{\boldsymbol{\beta}}_{i,t}$ is the post-ranking beta. $\mathbf{X}_{i,t-1}$ is the lagged vector of control variables. $\mathbf{ESG}_{i,t-1}$ represents the ESG scores available at $t-1$. $\gamma_{k,t}$ constitutes the coefficient of variable k in month t. Finally, $u_{i,t}$ is the regression residual.

Each time-variant coefficient can be characterized as the realization of a random variable. The expected value of the estimated coefficients $\overline{\hat{\gamma}}_k$ is given by the average over the t coefficients:

$$\overline{\hat{\gamma}}_k = \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_{k,t}. \tag{3}$$

Fama and MacBeth (1973) propose a simple t-test to evaluate the statistical significance of these coefficients. Their approach is designed to rule out intra-temporal dependencies along the cross-section. According to Petersen (2009), standard errors can be biased downwards if the independent variables and the residuals are correlated over time. In our data set, neither a company's market capitalization nor its book-to-market ratio is independently distributed over time. Furthermore, Dorfleitner, Halbritter, and Nguyen (2014) show that ESG scores exhibit a low temporal variability. The current rating is largely dependent on the rating of prior periods. In order to test for serial correlations in our panel data model, we perform a Wooldridge (2002) test. The t-statistics provide clear evidence for intertemporal dependencies, leading to biased standard errors of the traditional Fama and MacBeth (1973) approach (Petersen, 2009).

Hence, Fama and French (2002) suggest correcting the standard errors through the correlation of the estimated coefficients ($\hat{\rho}$). According to Petersen (2009) this leads to a substantial improvement of the standard errors. As a result, the test statistic is calculated as follows

$$t\left(\overline{\hat{\gamma}}_k\right) = \frac{\overline{\hat{\gamma}}_k}{\frac{\sigma_{\hat{\gamma}_k}}{\sqrt{T}}\sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}}. \tag{4}$$

As a robustness check of the cross-sectional approach, we additionally perform pooled OLS regressions with SIC-clustered standard errors. More than 500 SIC industry classes in our data set ensure a sufficient number of clusters to obtain an unbiased test statistic. While the Fama and MacBeth (1973) procedure equally weights each period, the pooled OLS method equally weights each observation. Therefore, both results may vary in the case of an unbalanced data set.

5. Results

5.1. ESG portfolios

Table 2 presents the regression results of the high and low portfolios as well as of the difference portfolios. Concerning the ASSET4 ratings, all high–low portfolios exhibit a positive alpha. While the high portfolios are largely in line with the overall market, the low portfolios slightly underperform the market. However, compared to earlier studies the abnormal returns are rather small and are therefore statistically insignificant. The portfolios based on the Bloomberg data set do not provide evidence for a significant relationship between ESG scores and returns, either. Only companies featuring a high social score generate significant positive abnormal returns of more than five percent per year. The high–low portfolio as related to corporate governance indeed shows negative alphas of up to two percent. The KLD portfolios do not offer significant return differences between high- and low-rated firms, either. Furthermore, the inclusion of controversial business involvement data has no significant effect on the alphas. Solely companies with a high environmental rating or a low corporate governance score achieve a significant outperformance up to almost three percent per year.

In summary, in all three samples the particular ESG ratings show a significantly lower influence on the financial performance than previous studies indicate. Although the alphas are not significant, we can still see differences between the three providers. In order to cancel out sample specific effects, we next restrict the analysis to companies providing ESG scores of all three agencies. Table 3 presents the estimated coefficients of the Carhart (1997) four-factor model. Concerning Bloomberg, the results are similar to the full sample model since portfolios are barely affected by the adjustment. ASSET4 still does not show a significant out- or underperformance of the high–low portfolios. In contrast, the KLD results are strongly influenced by the sample restriction. The alphas of the high–low portfolios are now negative. These findings point out that direction and magnitude of performance differences are not only dependent on the rating approach but also on the underlying company sample.

Regarding the factor loadings of the Carhart (1997) four-factor model, we can see notable differences between the betas of the high and low portfolios. Corporations with a higher ESG score are exposed to a lower systematic risk resulting in a lower beta. Some of the beta differences are even significant. Solely, the environmental and social factors of KLD lead to opposite implications. In addition to this, most high–low portfolios indicate a significantly different influence of the size factor between companies having a high or low score. This applies both to the overall ESG and the single pillar portfolios. For each of the dimensions, firms with a high score are therefore less strongly subjected to size risk. In the case of ASSET4 and Bloomberg, the HML coefficient is higher for companies with high ESG ratings. For some scores, the HML difference is even significant. This provides evidence that firms with high ESG scores also feature a higher book-to-market risk. The implications are contrary for the KLD sample. Beside the corporate governance rating, all high–low portfolios show significant negative loadings on the value factor. Firms with a high KLD rating also have lower weights on the momentum premium. For ASSET4 the results indicate an inverse relationship. The Bloomberg scores do not exhibit an obvious trend.

Subsequently, we consider a number of robustness checks to improve the validity of our results. Table 4 illustrates the alphas of the high–low portfolios obtained by applying the best-in-class approach. The findings do not suggest a link between the ESG score and the return. All alphas are insignificant while the magnitude and direction do not follow a certain pattern. Differences between the three data providers

Table 2
ESG portfolios: Time-series regressions based on full sample.

		Alpha	MKT	SMB	HML	WML	R^2
ASSET4							
ESG	High	−0.001	0.939***	−0.248***	0.048	0.027***	0.973
	Low	−0.011	1.042***	0.180***	0.011	−0.024	0.927
	High–low	0.011	−0.102	−0.428***	0.037	0.051	0.364
ENV	High	0.001	0.949***	−0.253***	0.046	0.014	0.969
	Low	0.000	1.072***	0.055	−0.099	−0.045**	0.932
	High–low	0.001	−0.122	−0.308***	0.144	0.059**	0.287
SOC	High	0.004	0.931***	−0.196***	0.036	0.002	0.972
	Low	−0.010	1.121***	0.126	−0.234**	0.010	0.903
	High–low	0.013	−0.190**	−0.322***	0.270**	−0.008	0.292
GOV	High	0.000	0.941***	−0.218***	0.012	0.039***	0.970
	Low	−0.006	1.063***	−0.015	−0.022	−0.103***	0.933
	High–low	0.006	−0.122	−0.203*	0.035	0.143***	0.341
ECN	High	0.010	0.947***	−0.172***	−0.027	−0.010	0.971
	Low	0.000	1.151***	0.180***	−0.057	−0.001	0.921
	High–low	0.011	−0.203***	−0.352***	0.030	−0.009	0.364
Bloomberg							
ESG	High	0.012	0.972***	−0.267***	0.062*	0.040***	0.969
	Low	−0.002	1.015***	0.076	−0.129**	0.015	0.917
	High–low	0.014	−0.043	−0.342***	0.191**	0.025	0.145
ENV	High	0.020	0.959***	−0.113	0.137	−0.055*	0.911
	Low	0.021	1.068***	−0.037	−0.005	−0.001	0.935
	High–low	−0.001	−0.109	−0.076	0.142	−0.054	0.028
SOC	High	0.052**	1.018***	0.225**	−0.055	−0.102***	0.922
	Low	0.010	1.070***	−0.129**	0.084	−0.020	0.955
	High–low	0.041	−0.052	0.354***	−0.140	−0.082*	0.092
GOV	High	−0.002	0.990***	−0.289***	0.082***	0.013	0.975
	Low	0.018	1.012***	−0.014	−0.188***	−0.059**	0.927
	High–low	−0.020	−0.022	−0.274***	0.271***	0.072	0.177
KLD							
ESG	High	0.017	0.956***	−0.213***	−0.057	−0.075***	0.890
	Low	0.005	1.005***	−0.226***	0.194***	0.044	0.880
	High–low	0.012	−0.049	0.013	−0.251***	−0.119***	0.121
ESG*	High	0.015	0.964***	−0.215***	−0.055	−0.066***	0.897
	Low	0.007	0.982***	−0.193***	0.215***	0.037	0.867
	High–low	0.008	−0.018	−0.022	−0.269***	−0.103***	0.122
ENV	High	0.029*	0.987***	−0.18***	−0.051	−0.056	0.802
	Low	0.006	0.961***	−0.296***	0.204***	0.039	0.894
	High–low	0.022	0.026	0.116*	−0.255***	−0.095***	0.142
SOC	High	0.007	1.029***	−0.226***	−0.058	−0.029	0.907
	Low	0.013	0.961***	−0.187***	0.159***	−0.002	0.863
	High–low	−0.006	0.068*	−0.039	−0.217**	−0.027	0.089
GOV	High	−0.003	0.933***	−0.036	0.181**	−0.030	0.679
	Low	0.018**	1.024***	−0.241***	0.014	−0.004	0.949
	High–low	−0.021	−0.092**	0.205**	0.167*	−0.026	0.059

This table presents the results of the Carhart (1997) four-factor model over the variable sample period from 1991 to 2012 on a monthly basis. The regressions are run individually for each ESG score and portfolio type using the full available data. The high (low) portfolio consists of the 20% best (worst) performing companies in terms of a particular ESG score. The high–low portfolio trades the high-rated companies long while the low-rated companies are traded short. All portfolios are weighted by the firms' market capitalization. Annualized alphas, factor loadings concerning size, value and momentum as well as adjusted R^2s are reported. The explanatory factors are provided by Asness and Frazzini (2013). The standard errors are estimated using the Newey and West (1987) procedure. ***, ** and * indicate a significance level of 1%, 5% and 10%.

are even more obvious. The implications are consistent to the standard model.

Furthermore, Table 5 presents Carhart (1997) four-factor model of the high–low portfolios dependent on portfolio cut-off and weighting. With a few exceptions, all alphas are insignificant. There is also no pattern concerning strength and sign of the abnormal returns. At this it makes no difference whether the portfolios are value or equally weighted. All in all, the models do not support evidence for a relationship between the social and financial performance.

According to previous studies there could be a link between ESG scores and financial returns in earlier years. Table 6 presents the annualized Carhart (1997) four-factor model alphas for the high–low portfolios estimated for various subperiods. Concerning KLD, we can see a notable downward movement of the abnormal returns. Beside the corporate governance score, all difference portfolios exhibit positive alphas during the first period from 1991 to 2001. Most of them are statistically significant. Consistently with previous studies investors are able to achieve annual abnormal returns of up to 6.6% following a long-short strategy. Over time the positive alphas slowly diminish and therefore lack statistical significance. From 2002 to 2006 they are only found in a range of 2.5% when using value weighted portfolios. Equally weighted portfolios lead to even lower abnormal returns which are close to zero. In 2007 to 2012 all alphas with the exception of corporate governance are negative, some of them even being significantly negative. In this case, the long-short investor would now lose up to six percent per year. The portfolios based on the corporate governance score do not follow any particular trend while the alphas are mostly negative. Portfolios based on ASSET4 ratings show a similar development. The high–low alphas slightly converge to zero and the significances decline also. The same holds true for the high–low portfolios based on Bloomberg scores.

Summarizing, the ESG portfolio strategy does not support a relationship between the CFP and the CSP of a company measured in terms of ESG scores. The Carhart (1997) four-factor model cannot show significant return differences between high- and low-rated firms for any of the portfolios. Comparing the three data providers, we can also find

Table 3
ESG portfolios: Time-series regressions based on an overlapping sample.

		Alpha	MKT	SMB	HML	WML	R^2
ASSET4							
ESG	High	0.005	0.946***	−0.245***	0.053	0.036***	0.963
	Low	−0.003	1.124***	0.221**	−0.043	0.013	0.906
	High–low	0.008	−0.178*	−0.466***	0.096	0.023	0.336
ENV	High	0.000	0.957***	−0.236***	0.033	0.020	0.963
	Low	0.001	1.063***	0.169***	−0.134**	−0.054***	0.948
	High–low	−0.001	−0.106	−0.406***	0.168*	0.073***	0.345
SOC	High	0.015	0.933***	−0.212***	0.041	0.021*	0.962
	Low	0.007	1.106***	0.130	−0.253***	0.005	0.894
	High–low	0.008	−0.174*	−0.342***	0.294**	0.016	0.248
GOV	High	0.008	0.931***	−0.237***	0.032	0.057***	0.948
	Low	0.006	1.128***	−0.065	0.022	−0.106***	0.942
	High–low	0.002	−0.197**	−0.172*	0.009	0.163***	0.440
ECN	High	0.024	0.952***	−0.13**	−0.004	−0.009	0.959
	Low	−0.015	1.138***	0.134	−0.082	−0.033	0.914
	High–low	0.039	−0.186**	−0.264**	0.078	0.024	0.233
Bloomberg							
ESG	High	0.009	0.971***	−0.258***	0.061*	0.045***	0.968
	Low	−0.014	1.076***	0.114	−0.120	0.008	0.911
	High–low	0.024	−0.105	−0.372***	0.182*	0.037	0.205
ENV	High	0.005	0.993***	−0.046	0.207**	−0.072*	0.914
	Low	0.027	1.065***	−0.027	−0.010	−0.004	0.932
	High–low	−0.021	−0.072	−0.019	0.217*	−0.068	0.037
SOC	High	0.042**	1.026***	0.169	−0.020	−0.111***	0.920
	Low	0.010	1.067***	−0.128**	0.105	−0.011	0.951
	High–low	0.032	−0.041	0.297**	−0.125	−0.100*	0.070
GOV	High	−0.005	0.978***	−0.289***	0.116***	0.022**	0.970
	Low	0.022	1.019***	−0.052	−0.165***	−0.025	0.933
	High–low	−0.026	−0.041	−0.237***	0.281***	0.047	0.163
KLD							
ESG	High	−0.012	0.938***	−0.114	0.065	−0.042*	0.929
	Low	0.025	1.048***	−0.194*	−0.074	0.141***	0.843
	High–low	−0.035	−0.110	0.080	0.139	−0.183***	0.081
ESG*	High	−0.011	0.933***	−0.110	0.073	−0.043*	0.929
	Low	0.031	1.043***	−0.219**	−0.077	0.14***	0.849
	High–low	−0.041	−0.110	0.109	0.149	−0.183***	0.092
ENV	High	0.008	0.956***	−0.123*	−0.014	−0.065**	0.931
	Low	0.036	1.067***	−0.277**	0.032	0.159***	0.855
	High–low	−0.027	−0.111	0.154	−0.046	−0.225***	0.105
SOC	High	−0.020	0.967***	−0.168**	0.137*	−0.049*	0.929
	Low	0.006	1.027***	−0.097	−0.084	0.101***	0.879
	High–low	−0.026	−0.060	−0.071	0.221	−0.150***	0.093
GOV	High	0.007	0.952***	−0.137***	0.042	0.017	0.945
	Low	0.010	1.014***	−0.175***	0.065	−0.006	0.951
	High–low	−0.002	−0.062*	0.038	−0.023	0.023	−0.004

This table presents the results of the Carhart (1997) four-factor model over the variable sample period from 1991 to 2012 on a monthly basis. The regressions are run individually for each ESG score and portfolio type using only observations available for all providers. The high (low) portfolio consists of the 20% best (worst) performing companies in terms of a particular ESG score. The high–low portfolio trades the high-rated companies long while the low-rated companies are traded short. All portfolios are weighted by the firms' market capitalization. Annualized alphas, factor loadings concerning size, value and momentum as well as adjusted R^2s are reported. The explanatory factors are provided by Asness and Frazzini (2013). The standard errors are estimated using the Newey and West (1987) procedure. ***, ** and * indicate a significance level of 1%, 5% and 10%.

notable differences in the direction and magnitude of the estimated coefficients. Different ESG concepts are therefore non-consistent in terms of return predictability. The robustness tests illustrate that these results hold various modifications of the standard model as related to portfolio construction and best-in-class ratings. Overall, the findings strongly argue against the existing evidence proposing abnormal returns of an ESG portfolio strategy (Derwall et al., 2005; Eccles et al., 2014; Kempf & Osthoff, 2007; Statman & Glushkov, 2009). Splitting the sample into several subperiods reveals the determinants for this. We find a significant decline in the explanatory power of ESG scores over the last decade. While in earlier year companies featuring high ESG scores significantly outperformed their low counterparts, as of 2012 not one of the three rating concepts appears to be able to anticipate abnormal returns.

Table 4
ESG portfolios: High–low alphas based on best-in-class ratings.

	ESG	ESG*	ENV	SOC	GOV	ECN
ASSET4	−0.016		0.012	−0.018	−0.023	−0.005
Bloomberg	−0.027		−0.024	−0.005	0.001	
KLD	0.017	0.005	0.028	0.026	0.002	

This table presents the annualized high–low alphas of the four-factor model over the variable sample period from 1991 to 2012 on a monthly basis. The regressions are run individually for each ESG score using the full available data. The high–low portfolio buys the 20% best performing companies in terms of a particular ESG best-in-class score while the worst 20% are traded short. All portfolios are weighted by market capitalization. The explanatory factors are provided by Asness and Frazzini (2013). The standard errors are estimated using the Newey and West (1987) procedure. ***, ** and * indicate a significance level of 1%, 5% and 10%.

5.2. Cross-sectional regressions

Even if the ESG portfolio approach does not provide evidence of a relationship between the social and financial performance, there might be a dependence in the cross-section. Using the full panel structure of our three databases allows us to obtain a more detailed

Table 5
ESG portfolios: High–low alphas dependent on cut-off and weighting approach.

	Market capitalization weighted					Equally weighted				
	1%	5%	10%	25%	50%	1%	5%	10%	25%	50%
ASSET4										
ESG	−0.079	0.010	0.001	0.013	0.014	−0.058	−0.031	−0.014	0.015	0.018*
ENV	−0.065	−0.059**	−0.017	0.009	0.007	−0.097	−0.043**	−0.008	0.011	0.013
SOC	0.011	0.009	0.006	0.011	−0.017	−0.049	−0.001	0.014	0.026***	0.010
GOV	0.010	0.010	−0.022	0.015	0.018	−0.009	−0.015	−0.023*	−0.001	0.010
ECN	−0.034	0.024	−0.019	0.028	−0.004	−0.008	−0.005	0.000	0.014	0.007
Bloomberg										
ESG	0.041	0.006	−0.011	0.015	−0.001	−0.083	−0.106***	−0.048**	−0.013	−0.019
ENV	−0.143	0.058	−0.019	−0.027	−0.002	0.018	0.064	0.033	0.003	−0.006
SOC	0.115	0.035	0.000	0.040	0.025	0.060	0.003	0.003	0.003	0.038*
GOV	−0.122	0.063	0.025	−0.032*	−0.032*	−0.037	0.017	0.009	−0.027**	−0.024
KLD										
ESG	0.014	0.004	0.005	0.010	−0.002	−0.041	−0.002	−0.008	0.003	−0.003
ESG*	0.016	−0.005	−0.006	0.006	−0.006	−0.030	0.007	−0.011	−0.002	−0.008
ENV	−0.028	0.016	0.039	0.022	0.002	−0.013	0.037	0.019	−0.003	−0.014
SOC	−0.074*	−0.027	−0.003	−0.008	−0.003	−0.071	−0.011	0.001	0.004	0.003
GOV	0.033	−0.010	−0.032	−0.032	0.007	0.012	−0.029	−0.034	−0.039*	−0.027***

This table presents the annualized high–low alphas of the Carhart (1997) four-factor model over the variable sample period from 1991 to 2012 on a monthly basis. The regressions are run individually for each ESG score using the full available data. The high–low portfolio buys the best performing companies in terms of a particular ESG score while the worst are traded short. Portfolio cut-offs of 1%, 5%, 10%, 25% and 50% are applied, both in a market capitalization and equally weighted version. The explanatory factors are provided by Asness and Frazzini (2013). The standard errors are estimated using the Newey and West procedure. ***, ** and * indicate a significance level of 1%, 5% and 10%.

understanding of how ESG scores can help in predicting returns. Compared to the portfolio approach, all companies and observations can be included in the analysis. Furthermore, there is no need for assumptions with regard to the portfolio construction.

Table 7 presents the estimated coefficients for the monthly Fama and MacBeth (1973) regressions. In terms of ASSET4 and Bloomberg, model (I) indicates a significant link between the overall ESG rating and the return. An increase of the total Bloomberg ESG score by one point implicates a monthly return increment of 0.014%. The total ESG variable of ASSET4 exhibits a significant coefficient of 0.008%. If we assume a 20 points higher ASSET4 ESG score which is relatively high on a scale between 0 and 100, the annualized return would ceteris paribus increase by 1.92%. For KLD both the overall ESG score and the ESG* score feature insignificant coefficients.

To identify the particular determinants of these results, we need to consider the individual pillars. Model (II) employs the subcriteria relating to environment, social and corporate governance in order to explain returns in the cross-section. Concerning the ASSET4 sample, the findings reveal that the significant influence of the overall ESG score is mainly driven by the economic rating. Increasing the economic score by one point leads to a return growth of 0.014%. The environmental and social pillars do not have significant explanatory power. The corporate governance score even has a slightly significant negative relationship to the financial performance. The Bloomberg sample shows that only the social score exhibits a small significant coefficient of 0.007% while the other pillars do not indicate an influence on a firm's return. The significance of the overall ESG rating is mainly driven by the sum of the environmental and social indicators. In the case of KLD

Table 6
ESG portfolios: High–low alphas within various subperiods.

	Market capitalization weighted				Equally weighted			
	Full Sample	1991 2001	2002 2006	2007 2012	Full Sample	1991 2001	2002 2006	2007 2012
ASSET4								
ESG	0.011		0.020	0.016	0.009		0.013	0.003
ENV	0.001		−0.007	0.012	0.009		−0.044	0.028
SOC	0.013		0.035	0.007	0.021**		0.024*	0.015
GOV	0.006		−0.009	−0.005	−0.021**		−0.033**	−0.025**
ECN	0.011		0.013	0.034	0.005		−0.020	0.006
Bloomberg								
ESG	0.014		0.015	0.008	−0.013		−0.015	−0.019
ENV	−0.001		−0.127	−0.013	0.016		−0.180***	0.005
SOC	0.041		−0.073	0.020	0.008		−0.146**	0.001
GOV	−0.020		−0.017	−0.020	−0.025		0.149*	−0.030
KLD								
ESG	0.012	0.054**	0.026	−0.002	−0.004	0.055***	0.003	−0.055*
ESG*	0.008	0.042*	0.023	−0.007	−0.010	0.046**	0.004	−0.060**
ENV	0.022	0.056	0.024	−0.003	0.003	0.066***	−0.024	−0.040
SOC	−0.006	0.019	0.027	−0.017	0.004	0.039*	0.004	−0.024
GOV	−0.021	−0.047	−0.043	0.014	−0.033	−0.035	−0.012	−0.045*

This table presents the annualized high–low alphas of the Carhart (1997) four-factor model over the subperiods 1991 to 2001, 2002 to 2006 and 2007 to 2012 as well as the full sample period. The regressions are run individually for each ESG score using the full available data. The high–low portfolio buys the 20% best performing companies in terms of a particular ESG score while the worst 20% are traded short. Both a market capitalization and equally weighted version is presented. The explanatory factors are provided by Asness and Frazzini (2013). The standard errors are estimated using the Newey and West (1987) procedure. ***, ** and * indicate a significance level of 1%, 5% and 10%.

Table 7
Modified Fama-MacBeth regressions based on the full sample.

	ASSET4		Bloomberg		KLD		
	(I)	(II)	(I)	(II)	(Ia)	(Ib)	(II)
BETA	−0.020	0.029	0.093	−0.001	0.070	0.064	0.038
	(−0.138)	(0.210)	(0.417)	(−0.007)	(0.459)	(0.424)	(0.246)
lnSIZE	−0.313***	−0.332***	−0.335***	−0.203***	−0.032	−0.030	−0.047
	(−4.454)	(−3.987)	(−5.711)	(−4.286)	(−1.285)	(−1.234)	(−1.641)
lnBM	0.280***	0.281***	0.284*	0.255**	0.524***	0.526***	0.530***
	(4.652)	(4.521)	(1.950)	(2.103)	(14.080)	(14.129)	(14.312)
MOM	0.113***	0.093***	0.061	0.023	0.083***	0.083***	0.085***
	(6.465)	(5.257)	(1.623)	(0.591)	(2.984)	(2.977)	(2.944)
ESG	0.008***		0.014***		0.007		
	(4.362)		(3.953)		(0.858)		
ESG*						0.009	
						(1.001)	
ENV		0.001		0.004			0.006***
		(1.498)		(0.729)			(2.615)
SOC		−0.002		0.007**			0.003
		(−1.009)		(2.271)			(0.522)
GOV		−0.003**		−0.002			−0.004***
		(−2.597)		(−0.224)			(−3.567)
ECN		0.014***					
		(10.649)					

This table presents the results of the adjusted Fama and MacBeth (1973) model over the variable sample period from 1992 to 2012 on a monthly basis. Model (I) only considers the overall ESG score while model (II) investigates the impact of the particular pillars. BETA, lnSIZE, lnBM and MOM are control variables with regard to beta, market capitalization, book-to-market ratio and average returns of the last 12 months. The dependent variable is given in percentage points. Standard errors are adjusted for autocorrelation. ***, ** and * indicate a significance level of 1%, 5% and 10%.

both the social and the corporate governance variables exhibit significant coefficients. Nevertheless, they feature the opposite direction and the effects cancel out in sum. In general, corporate governance practices appear to have a negative impact on a firm's return. This may be driven by the companies' size as large corporations tend to achieve higher governance ratings and lower returns than small firms (Bauer, Günster, & Otten, 2003; Humphrey, Lee, & Shen, 2012). Since we only control for the market capitalization, other aspects, such as the number of employees, could possibly contribute. The governance rating may proxy these effects.

In addition to the full sample, we consider an overlapping sample only including companies rated by all three providers. This allows us to compare the three rating concepts without being subjected to sample-specific effects. Table 8 presents the regression results. Model (I) finds a significant relationship between the ASSET4 and Bloomberg overall ESG scores and the returns. For each ESG point the monthly return increases by 0.016%. Concerning KLD, the overall ESG ratings now negatively influence the financial performance in the cross-section, even though not significantly. The particular pillars of KLD notably lose ground compared to the full sample. Solely the corporate governance score exhibits a small significant coefficient of 0.005%. The subcriteria of Bloomberg are fairly consistent with the standard model whereby only the social score is a significant determinant of returns. The ASSET4 coefficients are notably strengthened but the corporate

Table 8
Modified Fama-MacBeth regressions based on an overlapping sample.

	ASSET4		Bloomberg		KLD		
	(I)	(II)	(I)	(II)	(Ia)	(Ib)	(II)
BETA	0.052	0.076	0.095	0.051	0.092	0.094	0.160
	(0.215)	(0.284)	(0.403)	(0.224)	(0.393)	(0.402)	(0.759)
lnSIZE	−0.353***	−0.343***	−0.318***	−0.221***	−0.229***	−0.229***	−0.258***
	(−6.062)	(−6.028)	(−4.563)	(−3.818)	(−4.008)	(−3.978)	(−3.244)
lnBM	0.315**	0.329**	0.314**	0.265**	0.294**	0.295**	0.278**
	(2.231)	(2.318)	(2.120)	(1.979)	(2.165)	(2.160)	(2.178)
MOM	0.060	0.036	0.060	−0.014	0.052	0.052	0.045
	(1.447)	(0.908)	(1.391)	(−0.362)	(1.368)	(1.372)	(1.307)
ESG	0.009***		0.016***		−0.008		
	(8.478)		(4.162)		(−1.272)		
ESG*						−0.008	
						(−1.174)	
ENV		0.004**		0.006			−0.010
		(1.995)		(0.910)			(−1.164)
SOC		−0.009***		0.007**			−0.001
		(−2.932)		(2.563)			(−0.443)
GOV		0.005		0.002			0.005*
		(1.652)		(0.172)			(1.708)
ECN		0.017***					
		(10.303)					

This table presents the results of the adjusted Fama and MacBeth (1973) model over the variable sample period from 1992 to 2012 on a monthly basis using only observations available for all providers. Model (I) only considers the overall ESG score while model (II) investigates the impact of the particular pillars. BETA, lnSIZE, lnBM and MOM are control variables with regard to beta, market capitalization, book-to-market ratio and average returns of the last 12 months. The dependent variable is given in percentage points. Standard errors are adjusted for autocorrelation. ***, ** and * indicate a significance level of 1%, 5% and 10%.

Table 9
Modified Fama-MacBeth regressions for various subperiods.

	ASSET4		Bloomberg		KLD		
	(I)	(II)	(I)	(II)	(Ia)	(Ib)	(II)
Panel A: 2002 bis 2006							
BETA	0.090	0.087	0.004	0.158	−0.247	−0.249	−0.283
	(0.268)	(0.263)	(0.010)	(0.313)	(−0.662)	(−0.661)	(−0.728)
lnSIZE	−0.467***	−0.514***	−0.442***	−0.334**	−0.094	−0.095	−0.107*
	(−10.428)	(−9.11)	(−2.878)	(−2.518)	(−1.500)	(−1.530)	(−1.769)
lnBM	0.400***	0.404***	0.667***	0.610***	0.652***	0.653***	0.655***
	(7.386)	(6.844)	(5.403)	(7.631)	(13.427)	(13.395)	(13.139)
MOM	0.092***	0.077**	0.056	−0.009	0.027	0.027	0.029
	(2.937)	(2.542)	(0.398)	(−0.082)	(1.252)	(1.240)	(1.314)
ESG	0.009***		0.018***		−0.012		
	(5.093)		(2.926)		(−0.991)		
ESG*						−0.013	
						(−0.949)	
ENV		0.001		0.022**			−0.002
		(1.469)		(2.011)			(−0.389)
SOC		0.003***		0.014*			−0.007
		(2.913)		(1.906)			(−1.027)
GOV		−0.004***		−0.023**			−0.004***
		(−2.917)		(−2.121)			(−2.632)
ECN		0.010**					
		(2.641)					
Panel B: 2007 bis 2012							
BETA	−0.112	−0.019	0.128	−0.065	0.148	0.150	0.124
	(−0.446)	(−0.077)	(0.405)	(−0.209)	(0.553)	(0.561)	(0.486)
lnSIZE	−0.183***	−0.178***	−0.292***	−0.151**	0.039*	0.037*	0.047**
	(−3.774)	(−3.635)	(−4.139)	(−2.608)	(1.766)	(1.734)	(2.185)
lnBM	0.178***	0.176***	0.131	0.113	0.400***	0.400***	0.407***
	(4.340)	(4.623)	(1.519)	(1.476)	(6.968)	(6.935)	(6.795)
MOM	0.130***	0.106***	0.063	0.036	0.039	0.039	0.037
	(3.213)	(2.654)	(1.263)	(0.557)	(1.211)	(1.210)	(1.150)
ESG	0.006**		0.013***		−0.010		
	(2.198)		(2.890)		(−1.429)		
ESG*						−0.011	
						(−1.325)	
ENV		0.001		−0.002			0.004
		(0.498)		(−0.505)			(0.642)
SOC		−0.007***		0.004			−0.014***
		(−2.895)		(1.518)			(−4.806)
GOV		−0.002		0.007			−0.000
		(−0.829)		(1.358)			(−0.048)
ECN		0.018***					
		(7.554)					

This table presents the results of the adjusted Fama and MacBeth (1973) model over the subperiods from 2002 to 2006 (panel A) and from 2007 to 2012 (panel B). Model (I) only considers the overall ESG score while model (II) investigates the impact of the particular pillars. BETA, lnSIZE, lnBM and MOM are control variables with regard to beta, market capitalization, book-to-market ratio and average returns of the last 12 months. The dependent variable is given in percentage points. Standard errors are adjusted for autocorrelation. ***, ** and * indicate a significance level of 1%, 5% and 10%.

governance score loses statistical significance. Overall, the findings emphasize that the evidence in favor of a link between the social and financial performance is strongly dependent on the underlying sample and rating provider consistent to the results of Dorfleitner, Halbritter, and Nguyen (2014).

We also split the full sample in two subperiods, because in the context of the ESG portfolios the implications are conditional on the sample period. Table 9 presents the estimated coefficients for the Fama and MacBeth (1973) model separately for the years from 2002 to 2006 (panel A) and from 2007 to 2012 (panel B). Consistent with the ESG portfolios, model (I) suggests a slow decline in the explanatory power of the overall ESG variables. Nevertheless, the total ESG score of ASSET4 and Bloomberg still has significant power for explaining returns in the cross-section. Model (II) reveals the determinants of the total variables. The environmental and corporate governance scores of ASSET4 become less important whereas the economic performance gathers strength. In the case of Bloomberg, all ESG subcriteria lose their statistical significance from panel A to panel B. Except for the social score, this is also valid for the KLD data. Altogether, these results indicate a slightly declining impact of sustainability issues on the financial performance.

Summarizing the findings of the cross-sectional analysis, we find significant differences between the ESG concepts of the three providers. The overall ESG scores of ASSET4 and Bloomberg both have a significant influence on the returns. These results are robust for different subperiods. This indicates that there could be a relationship between ESG ratings and returns, even if investors are not able to exploit it through an ESG portfolio strategy. As opposed to this, the overall KLD scores do not provide evidence for a link between the ESG level and the financial performance. This is consistent with Galema et al. (2008) and Manescu (2011). Considering the individual pillars shows that the impact is mainly driven by only half of the ESG indicators. Aggregating the subcriteria leads to overlaying effects while some scores even cancel each other out. In terms of the particular pillars, we do not find consistent patterns among the three data sets. This emphasizes that the rating concept choice has a significant impact on the implications. A restriction of the sample to firms, which are covered by all three providers, leads to significant changes of some coefficients. Furthermore, the subperiods indicate a slowly diminishing explanatory power of ESG variables. As a robustness check we also estimate the models using pooled OLS regressions with SIC clustered standard errors. The results are fairly similar to those of the Fama and MacBeth (1973) procedure.

6. Conclusion

Applying two different approaches, this paper investigates the relationship between the corporate social and financial performance based on ESG ratings. Although previous empirical literature suggests a positive link between ESG rating levels and returns, we provide a critical review due to a number of concerns. Previous research identifies significant variations in the characteristics of several ESG rating concepts. As a consequence, it is crucial to address this question using more than a single ESG data set. Furthermore, most studies rely on data ending before 2007. Due to an enormous development of the SRI market over the last decade, a current sample is decisive. To gain evidence from different perspectives, we employ an ESG portfolio strategy as well as Fama and MacBeth's (1973) cross-sectional regressions.

The ESG portfolios do not show significant return differences between companies featuring high and low ESG rating levels. This applies both to the overall scores and to the particular pillars. This finding is robust for a variation of portfolio cut-offs as well as weightings. A best-in-class approach using sector-specific ESG scores does not generate abnormal returns either. These results strongly argue against previous studies suggesting abnormal returns of an ESG portfolio strategy (Derwall et al., 2005; Eccles et al., 2014; Kempf & Osthoff, 2007; Statman & Glushkov, 2009). Splitting the sample into three subperiods reveals the main determinant of this apparent contradiction. The Carhart (1997) four-factor model shows an obvious decline of the preceding outperformance over the last years.

Although ESG portfolios are not able to detect a link between the social and financial performance, the Fama and MacBeth (1973) regressions suggest an ambiguous significant influence of some ESG variables in the cross-section. Nevertheless, the influence is strongly dependent on the particular ESG rating provider. Furthermore, we do not identify a systematic pattern concerning the individual ESG dimensions among the three data sets. However most effects are robust for a number of subperiods, the results also provide evidence of a decreasing influence of ESG variables on the returns. This is consistent with the outcomes of the Carhart (1997) four-factor model.

In summary, this study strongly questions whether there is actually a relationship between ESG ratings and returns which is exploitable with a trading strategy in the sense of the Carhart (1997) four-factor model. This result is relevant both for researchers and for investors who focus on a portfolio composition based on ESG ratings. Even if recent evidence by Edmans (2011) or Dorfleitner, Utz, and Wimmer (2014) indicates that financial benefits of a strong CSP may only become visible if the respective stocks are held for a long time, social responsibility is considered worthwhile as a category of its own by many investors — even without yielding an additional return.

References

Asness, C.S., & Frazzini, A. (2013). The devil in HML's details. *Journal of Portfolio Management*, *39*(4), 49–68.

Bauer, R., Günster, N., & Otten, R. (2003). Empirical evidence on corporate governance. The effect on stock returns, firm value and performance. *Journal of Asset Management*, *5*(2), 91–104.

Bauer, R., Koedijk, K., & Otten, R. (2005). International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance*, *29*(7), 1751–1767.

Belghitar, Y., Clark, E., & Deshmukh, N. (2014). Does it pay to be ethical? Evidence from the FTSE4Good. *Journal of Banking and Finance*, *47*, 54–62.

Bello, Z.Y. (2005). Socially responsible investing and portfolio diversification. *Journal of Financial Research*, *28*(1), 41–57.

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

Breusch, T.S., & Pagan, A.R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287–1294.

Carhart, M.M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Derwall, J., Guenster, N., Bauer, R., & Koedijk, K. (2005). The eco-efficiency premium puzzle. *Financial Analysts Journal*, *61*(2), 51–63.

Dorfleitner, G., Halbritter, G., & Nguyen, M. (2014). Measuring the level and risk of corporate responsibility — An empirical comparison of different ESG rating approaches. *Working Paper*. Universität Regensburg.

Dorfleitner, G., Utz, S., & Wimmer, M. (2014). Patience pays off — Financial long-term benefits of sustainable management decisions. *Working Paper*. Universität Regensburg.

Durbin, J., & Watson, G.S. (1971). Testing for serial correlation in least squares regression iii. *Biometrika*, *58*(1), 1–19.

Eccles, R.G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, *60*(11), 2835–2857.

Edmans, A. (2011). Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics*, *101*(3), 621–640.

Fama, E.F. (1991). Efficient capital markets: II. *Journal of Finance*, *46*(5), 1575–1617.

Fama, E.F., & French, K.R. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*(2), 427–465.

Fama, E.F., & French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56.

Fama, E.F., & French, K.R. (2002). Testing trade-off and pecking order predictions about dividends and debt. *Review of Financial Studies*, *15*(1), 1–33.

Fama, E.F., & MacBeth, J.D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636.

Galema, R., Plantinga, A., & Scholtens, B. (2008). The stocks at stake: Return and risk in socially responsible investment. *Journal of Banking and Finance*, *32*(12), 2646–2654.

Hamilton, S., Jo, H., & Statman, M. (1993). Doing well while doing good? The investment performance of socially responsible mutual funds. *Financial Analysts Journal*, *49*(6), 62–66.

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, *93*(1), 15–36.

Humphrey, J.E., Lee, D.D., & Shen, Y. (2012). Does it cost to be sustainable? *Journal of Corporate Finance*, *18*(3), 626–639.

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*(5), 908–922.

Kreander, N., Gray, R., Power, D., & Sinclair, C. (2005). Evaluating the performance of ethical and non-ethical funds: A matched pair analysis. *Journal of Business Finance and Accounting*, *32*(7/8), 1465–1493.

Lee, D.D., & Faff, R.W. (2009). Corporate sustainability performance and idiosyncratic risk: A global perspective. *Financial Review*, *44*(2), 213–237.

Lee, D.D., Faff, R.W., & Rekker, S.A. (2013). Do high and low-ranked sustainability stocks perform differently? *International Journal of Accounting and Information Management*, *21*(2), 116–132.

Manescu, C. (2011). Stock returns in relation to environmental, social and governance performance: Mispricing or compensation for risk? *Sustainable Development*, *19*(2), 95–118.

Newey, W.K., & West, K.D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Orlitzky, M., Schmidt, F.L., & Rynes, S.L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, *24*(3), 403–441.

Petersen, M.A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, *22*(1), 435–480.

Sauer, D.A. (1997). The impact of social-responsibility screens on investment performance: Evidence from the Domini 400 social index and Domini equity mutual fund. *Review of Financial Economics*, *6*(2), 137–149.

Schröder, M. (2004). The performance of socially responsible investments: Investment funds and indices. *Financial Markets and Portfolio Management*, *18*(2), 122–142.

Schröder, M. (2007). Is there a difference? The performance characteristics of SRI equity indices. *Journal of Business Finance and Accounting*, *34*(1), 331–348.

Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal*, *56*(3), 30–39.

Statman, M. (2006). Socially responsible indexes: Composition, performance, and tracking error. *Journal of Portfolio Management*, *32*(3), 100–109.

Statman, M., & Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, *65*(4), 33–46.

USSIF (2012). Report on sustainable and responsible investing trends in the United States. *Forum for Sustainable and Responsible Investment*.

Utz, S., & Wimmer, M. (2014). Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. *Journal of Asset Management*, *15*(1), 72–82.

Wallis, M. v., & Klein, C. (2014). Ethical requirement and financial interest: A literature review on socially responsible investing. *Business Research* (http://dx.doi.org/10.1007/s40685-014-0015-7, in press).

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data*. MIT Press.