Financial Analysts Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ufaj20

ESG Rating Disagreement and Stock Returns

Rajna Gibson Brandon, Philipp Krueger & Peter Steffen Schmidt

To cite this article: Rajna Gibson Brandon, Philipp Krueger & Peter Steffen Schmidt (2021) ESG Rating Disagreement and Stock Returns, Financial Analysts Journal, 77:4, 104-127, DOI: 10.1080/0015198X.2021.1963186

To link to this article: https://doi.org/10.1080/0015198X.2021.1963186

Published online: 23 Sep 2021.

Financial Analysts Journal | A Publication of CFA Institute
https://doi.org/10.1080/0015198X.2021.1963186

ESG Rating Disagreement and Stock Returns

Rajna Gibson Brandon, Philipp Krueger, and Peter Steffen Schmidt

Rajna Gibson Brandon is a professor of finance at the Geneva School of Economics Management and Geneva Finance Research Institute, University of Geneva, Switzerland, and a research fellow at the European Corporate Governance Institute, Geneva, Switzerland. Philipp Krueger is an associate professor of responsible finance at the Geneva School of Economics Management and Geneva Finance Research Institute, University of Geneva, Switzerland, and senior chair at the Swiss Finance Institute, Geneva, Switzerland. Peter Steffen Schmidt is a senior research associate at the Department of Business Administration, University of Zurich, Switzerland.

Using environmental, social, and governance (ESG) ratings from seven different data providers for a sample of firms in the S&P 500 Index between 2010 and 2017, we studied the relationship between ESG rating disagreement and stock returns. We found that stock returns are positively related to ESG rating disagreement, suggesting a risk premium for firms with higher ESG rating disagreement. The relationship is primarily driven by disagreement about the environmental dimension. We discuss the practical implications of our findings for firms' equity cost of capital as well as for investment managers and asset owners who use ESG investment strategies.

Disclosure: The authors report no conflicts of interest.

PL Credits: 2.0

> There's so much disagreement about investing, and it's because nobody really knows.
>
> —*Robert J. Shiller*

Environmental, social, and governance (ESG) ratings nowadays feature prominently in the financial press, regulatory and policy debates, and academic studies, and they are also a hot topic in investment practice. Such ratings also increasingly shape investment decisions of institutional investors representing trillions of dollars in assets under management (Gibson, Glossner, Krueger, Matos, and Steffen 2021; GSIA 2016; USSIF 2020; PRI 2018). Recently, a lot of attention has been paid to divergence of ratings for the same firm issued by different ESG rating providers. For example, in a *Wall Street Journal* article, Mackintosh (2018) pointed out that Tesla was rated highly by MSCI regarding environmental issues in 2018. In contrast, FTSE reached the opposite conclusion, rating Tesla poorly on environmental matters. Other news outlets, policy-oriented think tanks, and practitioner-oriented publications have made similar observations (Wigglesworth 2018; Doyle 2018; Matos 2020).

In this article, we pursue two objectives: First, we report our systematic analysis of the level of disagreement about a firm's ESG rating to gain a better understanding of its magnitude and to determine whether ESG rating disagreement correlates with a firm's financial and accounting characteristics. Second, we consider whether diverging ESG ratings have real consequences for firms and investors by examining whether stock returns are related to ESG rating disagreement. Although contemporaneous work has attempted to explain the important question of "why" ESG ratings disagree (e.g., Berg, Koelbel, and Rigobon 2020; Christensen, Serafeim, and Sikochi, forthcoming), this study is a first attempt at examining the fundamental issue of whether ESG rating disagreement affects stock returns. Indeed, because of the prominent role that the equity cost of capital plays when financial analysts value firms and when chief financial officers (CFOs) decide how to allocate capital expenditures, a study of the impact of ESG rating disagreement on stock returns is likely to be relevant to them. We furthermore discuss how ESG rating disagreement could affect the performance of

equity investors who pursue sustainable investment strategies.

For the purpose of this study, we collected and studied ESG ratings from seven prominent ESG rating providers for firms in the S&P 500 Index between 2010 and 2017. We used data from Asset4 (now Refinitiv ESG), Sustainalytics (now Morningstar), Inrate, Bloomberg, FTSE, KLD (now MSCI), and MSCI IVA. We believe that our work has the most comprehensive data coverage among current papers that have studied the issue of ESG rating disagreement. For instance, Christensen et al. (forthcoming) used data from only three rating providers, and Berg et al. (2020) did not focus on the time-series dimension. In contrast, we built a panel dataset based on seven different ESG ratings. Note also that the focus of the other papers is very different from ours, in that they primarily examine *why* ESG ratings disagree.

We started our analysis by documenting some basic empirical facts about ESG rating disagreement in our sample of S&P 500 firms. We found, for example, that the average pairwise correlation between the ESG ratings of the seven rating providers is about 0.45. Surprisingly, the average pairwise correlation is lowest for the governance dimension (0.16) and highest for the environmental dimension (0.46). Our analysis of pairwise correlations between ESG ratings from different providers also highlights more subtle patterns in disagreement that go against the common beliefs about generalized ESG rating disagreement. For example, our analysis also shows that correlations between ratings from some providers are markedly high, with the maximum correlation being about 0.75.

We next studied whether disagreement varies in relation to observable firm-level financial and accounting characteristics. We provide evidence that disagreement tends to be more prevalent for the largest firms in the S&P 500 (perhaps because of the complexity of such firms) and for firms that do not have credit ratings (potentially because the information environment for these firms is of lower quality than for other firms). In contrast, more profitable firms tend to have lower ESG rating disagreement (perhaps because they are able to dedicate more resources to ESG policies and to their ESG disclosures). We also found that ESG rating disagreement is orthogonal to disagreement in the EPS forecasts issued by analysts. We further show that rating disagreement is generally more pronounced for firms that belong to the consumer durables and telecommunications industries, which provides important insights for financial analysts who cover firms in these sectors.

In our main empirical analysis, we measured ESG rating disagreement as the standard deviation of the available ESG ratings from the seven different data providers for a given firm at a given point in time. We calculated the disagreement measure for the total ESG rating and separately for the E, S, and G dimensions (or "pillars"). We then related monthly stock returns to our proxy of ESG rating disagreement while controlling for standard stock characteristics that are known to have predictive power in the cross-section of stock returns (e.g., size, momentum, and quality).

We found that stock returns are positively related to ESG rating disagreement. Further tests showed that the relationship is driven mainly by disagreement about the *environmental* rating. In terms of economic magnitude, we estimated that an interquartile range increase in ESG rating disagreement is associated with an increase of 92 bps in the annual cost of equity capital. Hence, ignoring differences in ESG rating disagreement in corporate valuation could lead to sizable mistakes when estimating the value of a firm's equity: Assume, for instance, a firm has perpetual annual free cash flow to equity (FCFE) of $100 million per year, an expected growth rate of the FCFE of g = 0%, and a cost of equity capital of r_E = 5%. An interquartile range increase in ESG rating disagreement combined with our empirical estimates would imply that the true cost of equity capital is 5.92%, suggesting an overestimation of the value of the firm's equity by about $311 million (or 18%) if the ESG rating disagreement is ignored.[1]

Our finding of a positive relationship between ESG rating disagreement and stock returns can be rationalized with a standard asset pricing argument: Higher ESG rating disagreement may be perceived as a source of uncertainty—in the spirit of Knightian uncertainty (that is, a lack of any quantifiable knowledge about some possible occurrence)—that commands an uncertainty premium. This explanation would clarify why uncertainty-averse investors taking on such additional exposure wish to be compensated by higher expected returns. Consistent with this conjecture, we found in standard portfolio sorts that a portfolio that was long stocks with a high level of disagreement and short stocks with a low level of disagreement generated monthly returns of about 21 bps (2.52% on an annual basis) for disagreement about the overall ESG and the environmental rating. We found similar magnitudes when adjusting

for well-known risk factors by using standard asset pricing models such as the Carhart (1997) or the Fama and French (2015) five-factor models.

Although many studies have examined the relationship between stock returns and average ESG ratings (see, e.g., Friede, Busch, and Bassen 2015 and references therein), ours is the first to systematically test whether the second moment of ESG ratings (i.e., disagreement about a firm's ESG performance) has consequences for stock returns. Another important contribution of our article is to move beyond simply documenting that disagreement about nonfinancial information exists and instead shedding light on whether such ESG disagreement has real consequences for firms, analysts, and investors. We also contribute importantly to the debate on why ESG rating disagreement exists by providing evidence of the firms' financial and accounting characteristics that correlate with ESG rating disagreement and by identifying those industry sectors that are most prone to ESG rating disagreement.

Overall, our empirical results should help financial analysts, academics, institutional investors, financial advisers, policymakers, regulators, and ultimately, firms themselves to better understand that, in addition to the sustainability performance as captured by average ESG ratings, the dispersion of these ratings also can have an economically meaningful impact on stock returns and thus on firms' equity cost of capital.

Literature Review

The use of ESG ratings in investment practice increased considerably during the last two decades and has skyrocketed recently. In parallel, ESG ratings are now also commonly used in economics, management, and finance research.[2] Given the complexity of measuring a firm's nonfinancial or ESG performance, the validity and convergence of these ratings have been debated critically in the management literature.[3] Chatterji, Durand, Levine, and Touboul (2016), for instance, studied the convergence of corporate social responsibility (CSR) ratings produced by six well-established information intermediaries. They documented a lack of agreement among information intermediaries that comes mainly from two sources—the absence of a common *theorization* and lack of *commensurability*.[4] These findings point out that firms' and professional investors' sustainable financing and investment decisions are potentially tainted by the choice of their rating providers. The same applies to the conclusions stemming from existing empirical studies conducted by academics.

Given the heightened concern that diverging ESG ratings have recently generated in both practitioner circles and the financial press, the topic has also spurred significant academic interest. Most of the contemporaneous work aims primarily at explaining the *why*—that is, what drives ESG rating disagreement. For instance, Christensen et al. (forthcoming) focused on the role of disclosure as a determinant of ESG rating disagreement and found that more disclosure leads to higher disagreement.[5] In addition, they pointed out that the relationship between a firm's average ESG rating and ESG rating disagreement is nonlinear. An important difference between their study and ours is that they focused primarily on explaining why disagreement exists, whereas we are more interested in examining the *consequences* arising from ESG rating disagreement and, specifically, whether ESG rating disagreement has implications for stock returns, a firm's cost of capital, and an investor's performance when investing sustainably.

Berg et al. (2020) also explored why ESG rating disagreement exists. They pursued a more granular approach, proposing a decomposition of the sources of ESG rating disagreement. By subdividing the ESG ratings of six providers into finer categories, Berg et al. identified three sources of ESG rating divergence. First, they highlighted that raters use different categories, which can lead to disagreement. They referred to this source as "scope divergence." Second, they pointed out that ESG raters measure identical categories differently, which they referred to as "measurement divergence." Finally, they highlighted "weight divergence," which results from raters attaching different weights to the different categories when generating an aggregated ESG rating. The authors found that most of the differences can be traced to measurement and scope divergence; weight divergence seems to play a minor role. In addition, Berg et al. found a "rater effect"—that is, ratings of one provider were positively correlated across different categories. The most important difference between Christensen et al. and Berg et al., on the one hand, and our study, on the other hand, is that their studies focused mainly on explaining why ratings disagree whereas we focus on whether there are consequences from ESG ratings' disagreements in the form of measurable effects on stock returns.

In a recent paper, Avramov, Cheng, Lioui, and Tarelli (2021) examined the relationships among stock returns, ESG ratings, and ESG rating disagreement.

Using a theoretical model that highlights the interplay between the average ESG rating and ESG rating disagreement, Avramov et al. showed that the average ESG rating is negatively associated with future stock performance only for low-ESG-disagreement stocks. In contrast to Avramov et al., we focus on ESG rating disagreement only.

Studying disagreement in ESG ratings is also reminiscent of the rich literature on heterogeneous beliefs in financial markets (see the discussion in the subsection "Possible Theoretical Explanations"). Many empirical studies have tested the relationship between dispersion in beliefs and stock returns in a variety of settings. These studies typically used the dispersion in analyst earnings forecasts as a proxy for the extent to which a stock is subject to heterogeneous beliefs. For example, Diether, Malloy, and Scherbina (2002) documented a significant and negative relationship between heterogeneous beliefs and stock returns. Anderson, Ghysels, and Juergens (2005) reached the opposite conclusion. They argued that disagreement about expected EPS is an additional priced risk factor and provided supportive empirical evidence for the pricing of this additional source of risk.

In the subsection "Possible Theoretical Explanations," we discuss another stream of the finance literature that would allow rationalizing the excess return for stocks with high ESG rating disagreement—namely, that these excess returns are a compensation for ESG information uncertainty (see also Viale, Garcia-Feijoo, and Giannetti 2014).

Data

To test how stock returns are related to ESG rating disagreement, we constructed a representative and homogeneous sample with available ESG ratings over the longest possible period. We faced the challenge that the availability of ESG data is restricted in both the cross-section and the time series. This specific limitation is not unique to our setting but applies generally to research concerned with ESG. To use a sample as homogeneous as possible and to maximize the number of available ESG ratings per firm (as well as the time-series dimension of the panel), we restricted our study to firms belonging to the S&P 500 and considered a sample period from 2010 through 2017.

We used financial data from CRSP and accounting data from Standard & Poor's Compustat.

We collected data from seven ESG data providers: (1) Asset4 (now Refinitiv),[6] (2) Sustainalytics,[7] (3) Inrate, (4) Bloomberg, (5) FTSE, (6) KLD,[8] and (7) MSCI Intangible Value Assessment (IVA).[9] According to a survey by Wong, Brackley, and Petroy (2019), the most important and commonly used providers are Sustainalytics, MSCI, Bloomberg, and Asset4. Later in the article, we report robustness analyses in which we restricted disagreement to the ratings from these four providers. In Appendix A, we provide further information about sample selection, dataset matching, variable definitions (in **Table A1**), and summary statistics of the variables used in this study (in **Table A2**).

Table 1 displays important features of these seven data providers. As shown in the "Origin" column, three providers are US-based (Bloomberg, KLD, and MSCI IVA), whereas two providers have their origins in Switzerland (Asset4 and Inrate). The other two providers can be traced back to origins in the Netherlands (Sustainalytics) and in the United Kingdom (FTSE).

In column 2 of Table 1, we show the rating scales used by each provider. For example, three providers apply a scale from 0 to 100 for their assessments; Inrate uses a scale of 1 to 12, which is based on sustainability assessments ranging from D– to A+. Originally, MSCI KLD did not provide a genuine scale itself. However, many academic studies (e.g., Lins, Servaes, and Tamayo 2017) have summed up KLD's judgment as to "strengths" and "concerns" separately and scaled both by the total number of strengths and concerns available. This course of action results in a scale of –1 to +1. Note that KLD also has strengths and concerns items for norms-oriented categories related to alcohol, military, firearms, gambling, nuclear, and tobacco, which we decided to ignore.

Because the rating scales differ not only in terms of their statistical support but also in terms of the distribution across the statistical support, a simple rescaling would not suffice to make the different ratings comparable. Therefore, we did the following to achieve comparability across rating providers: At each point in time, we sorted all stocks according to the ratings of the respective providers. We then calculated the individual rating-specific percentile ranks and used these as adjusted scores. Using ranked measures is also consistent with investment practice, in which investors compare the ranked value of a given signal relative to the ranked values of the signals for other firms. In the case of ties,

Table 1. ESG Data Providers

Data Provider	Origin (1)	Rating Scale (2)	Period Covered (3)	Number of Stocks (sample) (4)	Pillars (5)	Rating Style (6)
Asset4	Switzerland	0–100	Jan. 2010–Dec. 2017	438	E, S, G, Total	Disclosure oriented
Sustainalytics	Netherlands	0–100	Jan. 2010–Dec. 2017	459	E, S, G, Total	Best in class
Inrate	Switzerland	1–12	Jan. 2013–Dec. 2017	434	E, L, S, G, Total	Absolute ESG rating
Bloomberg	United States	0–100	Jan. 2010–Dec. 2017	463	E, S, G, Total	Disclosure oriented
FTSE	United Kingdom	0–5	Oct. 2014–Dec. 2017	442	E, S, G, Total	Best in class
MSCI KLD	United States	–1 to +1	Jan. 2010–Dec. 2017	468	E, S, G, Total	Absolute ESG rating
MSCI IVA	United States	0–10	Jan. 2010–Dec. 2017	456	E, S, G, Total	Best in class

Note: The data dimensions (e.g., environmental, social, and governance) are referred to as "pillars." The "L" here stands for "labor."

we assigned each company the average rank. We normalized these ranks between 0 and 1.

Column 4 of Table 1 shows the average number of sample stocks per year for which we observed an ESG rating from a given data provider. Sustainalytics, MSCI KLD, MSCI IVA, and Bloomberg had, on average, the best coverage (about 460 stocks). Inrate, Asset4, and FTSE had the least number of stocks, on average, with 434, 438, and 442, respectively. The average number of stocks for all providers is rather high—well above 400—and we therefore considered the sample to be representative for S&P 500 companies. Note, however, that the Inrate and FTSE ratings were available only for a subperiod of the overall sample period. Therefore, we report in column 3 the period for which data were available from a given provider.

The fifth column of Table 1 reports the pillar scores supplied by the providers as well as their rating styles. All providers supplied a total ESG score, an environmental score, a social score, and a (corporate) governance score. In addition, Inrate provided a labor score. Because the labor score captures a social topic, we used the average of the original social score and the labor score as the social score. In column 6, we highlight the rating styles used by these various data vendors, which may partially explain disagreements. For instance, MSCI KLD and Inrate provide an absolute ESG rating, whereas providers such as Sustainalytics and MSCI provide best-in-class ratings. Also, some rating providers (e.g., Bloomberg) are more geared toward capturing ESG disclosure quality than are other providers.

Analysis

In this main section of our article, we begin with descriptive statistics and correlations, consider the financial and accounting determinants of ESG rating disagreement, and then turn to the main research question: the relationship between stock returns and ESG rating disagreement. Next, we evaluate whether ESG rating disagreement provide a profitable signal for investing by reporting the results for portfolios based on ESG rating disagreement. We then consider the possible theoretical explanations for the relationship between ESG rating disagreement and stock returns, and finally, we discuss whether ESG rating disagreement affects standard equity risk measures.

Descriptive Statistics and Correlations.

Table 2 provides summary statistics and Pearson correlations between the ESG ratings from the seven different data providers. We show the results for the total rating and the E, S, and G pillars in separate panels. The first three columns display descriptive statistics for the ranked ESG scores from the

Table 2. Descriptive Statistics and Correlations, January 2010–December 2017

Data Provider	*N* (1)	Mean (2)	StD. EPS (3)	Pearson Correlations: Asset4 (4)	Sust. (5)	Inrate (6)	Bloom. (7)	FTSE (8)	KLD (9)
A. Total pillar									
Asset4	42,087	0.501	0.289						
Sustainalytics	44,078	0.501	0.289	0.752					
Inrate	26,037	0.501	0.284	0.233	0.303				
Bloomberg	44,464	0.501	0.289	0.750	0.693	0.124			
FTSE	17,220	0.501	0.288	0.568	0.614	0.267	0.586		
KLD	44,951	0.501	0.288	0.524	0.559	0.292	0.477	0.488	
MSCI IVA	43,775	0.501	0.289	0.396	0.434	0.318	0.303	0.266	0.439
Average correlation						0.447			
B. Environmental pillar									
Asset4	42,019	0.501	0.289						
Sustainalytics	44,020	0.501	0.289	0.706					
Inrate	26,036	0.501	0.286	0.305	0.487				
Bloomberg	37,624	0.501	0.289	0.647	0.557	0.206			
FTSE	17,220	0.501	0.288	0.654	0.678	0.368	0.607		
KLD	44,669	0.501	0.280	0.575	0.609	0.422	0.431	0.581	
MSCI IVA	43,580	0.501	0.289	0.233	0.352	0.404	0.187	0.239	0.312
Average correlation						0.455			
C. Social pillar									
Asset4	42,087	0.501	0.289						
Sustainalytics	44,078	0.501	0.289	0.617					
Inrate	26,037	0.501	0.288	0.133	0.143				
Bloomberg	44,364	0.501	0.288	0.685	0.527	0.062			
FTSE	17,220	0.501	0.288	0.637	0.501	0.106	0.560		
KLD	44,951	0.501	0.288	0.367	0.391	0.129	0.276	0.271	
MSCI IVA	43,775	0.501	0.289	0.266	0.303	0.236	0.202	0.191	0.337
Average correlation						0.330			
D. Governance pillar									
Asset4	42,087	0.501	0.289						
Sustainalytics	44,078	0.501	0.289	0.331					
Inrate	26,037	0.501	0.283	0.297	0.401				
Bloomberg	44,464	0.501	0.282	0.432	0.327	0.344			
FTSE	17,220	0.501	0.288	0.027	0.160	−0.029	−0.027		
KLD	44,951	0.501	0.248	0.104	0.089	0.081	0.153	−0.065	
MSCI IVA	43,775	0.501	0.288	0.132	0.135	0.145	0.060	0.023	0.133
Average correlation						0.155			

Note: We display in the last row of each panel the average pairwise correlation between providers. "StD. EPS" = disagreement about EPS.

different providers. The subsequent columns display the pairwise cross-correlations. We also show in the last row of each panel the average pairwise correlation between providers, calculated as the mean of the respective pairwise cross-correlations (separately for the total rating and the E, S, and G pillars).[10]

Note, first, that the average pairwise correlation for the overall ESG ratings in Table 2 is 0.45, which is much lower than average correlations between credit ratings issued by Moody's Investors Service and Standard & Poor's. According to Berg et al. (2020), correlations between those two credit-rating providers exceed 0.99. A point worth mentioning is that despite the commonly held belief of generalized ESG rating disagreement, the analysis of pairwise correlations in Table 2 also highlights more subtle patterns in ESG disagreement, in that the pairwise correlations between ratings from some providers can be relatively high. For instance, the correlations between the total rating of Asset4 and Sustainalytics and Asset4 and Bloomberg are about 0.75, and that between Sustainalytics and Bloomberg is about 0.69. This pattern may result from their similar rating styles, as documented in Table 1.

Next, we separately examine the average pairwise correlations between providers for the E, S, and G pillars (Panels B–D). Apart from the environmental pillar, these average pairwise correlations are generally lower than for the total rating, which probably results from discrepancies in aggregation and weighting procedures across the three pillars. Surprisingly, the average correlation is lowest for the governance pillar (0.16) and highest for the environmental pillar (0.46), but this result can also be rationalized: Environmental issues can be increasingly measured and quantified (e.g., water usage, greenhouse gas emissions), but the criteria applied to *quantifying* governance may likely differ among rating providers. In a similar spirit, the social rating is also likely to require more value judgments and is thus inherently more subjective than the environmental pillar, suggesting more disagreement among raters (i.e., lower correlations).

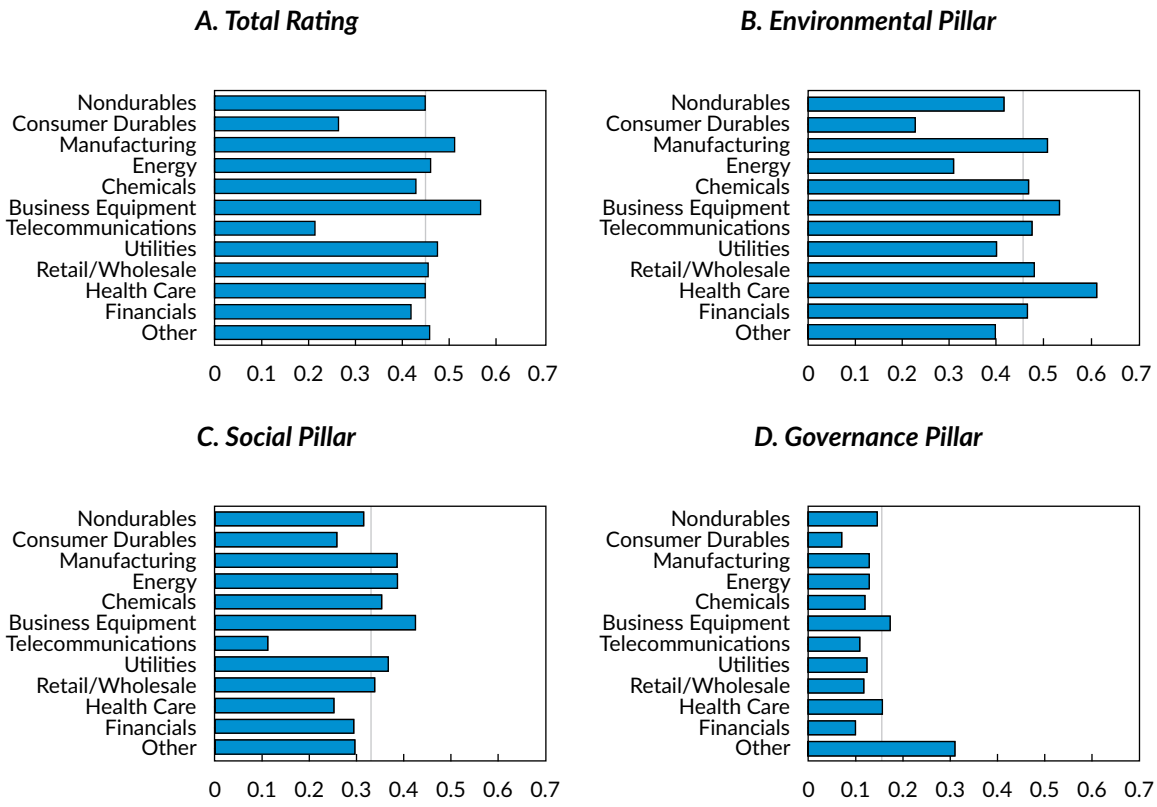
Our main argument for stating that the E rating is more objective and measurable than the others is this: We believe that not only is there more agreement on the issues that are important in the environmental dimension but also there are more systematic, regulation-driven, firm-level attempts to quantify these dimensions. For instance, a consensus now exists that greenhouse gas (GHG) emissions are an important dimension of a firm's environmental performance, and GHG emissions are increasingly measured. In a similar spirit, nowadays firms quantify water and electricity use. Although measurement of emissions is certainly not without problems (scaling of emissions, missing emissions data/imputation, voluntary reporting), firms at least have some basic guidance about how to measure them (see, for instance, the Greenhouse Gas Protocol[11]).[12] In contrast, we think that when it comes to S and G, there is (1) less agreement among providers on what the most important issues are and (2) a worse understanding of how to quantify these issues.

In **Figure 1**, we can see whether the average pairwise correlations between ESG ratings vary at the industry level. We plotted the average correlations across the seven ESG rating providers for each of the 12 Fama and French industries.[13] Some industry heterogeneity appears when it comes to correlations between ESG ratings. Average correlations in the total ratings (Panel A) are lowest in the consumer durables and telecommunications sectors. The low average pairwise correlation in the total rating for the consumer durables industry seems to be driven by the low pairwise correlations in the environmental ratings (Panel B) and governance ratings (Panel D). The low average correlation of the telecommunications sector results from the low correlations of the social ratings (Panel C). In contrast, ESG data providers seem to disagree the least (i.e., exhibit high average correlations) as to the total rating (Panel A) in the business equipment and manufacturing sectors. Another interesting observation is that rating providers also seem to disagree quite strongly (i.e., correlations are low) about governance in the finance sector. These findings on industry-variation in firms' ESG ratings could help industry-focused financial analysts distinguish their assessments and comparisons of firms, in that analysts are often confronted with metrics coming from various data providers.

Determinants of ESG Rating Disagreement.

We do not see our main contribution as studying determinants of ESG rating disagreement, but the comprehensiveness of our sample (in terms of rating providers covered) as well as the representativeness of the sample firms (S&P 500 firms) allow us to contribute also to this literature. Hence, to add to the existing literature that traces the origins to disagreement in scopes, measurement, rating methodologies, and ESG firms' disclosure policies, we examined whether ESG rating disagreement also correlates with observable firm-level financial and accounting characteristics. We used the standard deviation of ESG ratings available

Figure 1. Average Correlations by Fama and French 12-Industry Classification



Notes: Shown are average pairwise Pearson correlations between the normalized ratings of the ESG data providers for each of the 12 Fama and French industries. The vertical line represents the average correlation across all industries.

for a given firm at a given point in time as the dependent variable. We calculated this measure for the total rating and also separately for the E, S, and G pillars. We explored how the levels of rating disagreement correlate with variables falling in one of the following five categories: (1) balance sheet–related data, (2) industry-related data, (3) investor transparency, (4) valuation, and (5) price.

We used pooled panel regressions in which the rating disagreement measures served as dependent variables. We also included industry-month fixed effects. Standard errors were double clustered at the firm and month levels. **Table 3** displays the regression results for disagreement about the total ratings and separate results for disagreement about the pillars.

Essentially, Table 3 shows that three financial variables play a role in explaining rating disagreements. First, the more profitable firms (in terms of gross profitability) are subject to lower ESG rating disagreement (see columns 1 and 2). Second, firms without a credit rating exhibit higher ESG rating disagreement (see columns 1 and 3), as do larger firms (see columns 3 and 4).[14] These results seem intuitive: Profitable firms may be viewed less critically by ESG analysts, perhaps because they have more resources available to shape and disclose their ESG policies. In contrast, firms without a credit rating are subject to a less transparent information environment, making their assessment in terms of ESG more difficult. Also, large firms might be more diversified and complex and are also analyzed more thoroughly by ESG data providers, explaining why they exhibit higher rating disagreement. Finally, note that ESG rating disagreement is orthogonal to disagreement about EPS (StD. EPS in Table 3).

In addition, three other variables seem to affect disagreement about individual pillars of the ESG rating. For instance, tangibility plays a specific role, in that firms with more tangible assets tend to have lower disagreement in their environmental ratings. Again, this result seems intuitive, because firms with more tangible assets are also likely to have more negative impacts on the environment (e.g., higher GHG

Table 3. Determinants of ESG Rating Disagreement, January 2010–December 2017 (*t*-statistics in parentheses)

	Dependent Variable: ESG Rating Disagreement			
Pillars	Total (1)	Environmental (2)	Social (3)	Governance (4)
Balance sheet related				
Tangibility	−0.013	**−0.030**	−0.010	−0.011
	(−0.909)	**(−2.027)**	(−0.673)	(−0.859)
Current ratio	0.012	−0.001	−0.001	0.010
	(1.064)	(−0.081)	(−0.114)	(1.230)
Leverage	−0.015	−0.012	0.006	−0.005
	(−1.464)	(−1.163)	(0.657)	(−0.594)
Gross profitability	−0.028	−0.027	−0.021	−0.010
	(−1.954)	(−1.955)	(−1.610)	(−0.786)
Industry				
HHI	**0.029**	0.021	0.012	−0.008
	(2.183)	(1.593)	(0.911)	(−0.775)
Multisegment	0.001	−0.002	0.002	0.002
	(0.085)	(−0.293)	(0.371)	(0.416)
Investor transparency				
No credit rating	0.015	−0.010	**0.020**	−0.012
	(1.678)	(−1.208)	**(2.713)**	(−1.851)
Institutional ownership	0.014	0.015	0.010	−0.004
	(1.491)	(1.794)	(1.073)	(−0.551)
Number of analysts	−0.008	−0.011	−0.004	0.000
	(−0.906)	(−1.183)	(−0.446)	(0.035)
StD. EPS	0.005	0.006	0.013	0.005
	(0.581)	(0.887)	(1.579)	(0.646)
Valuation				
Book-to-market ratio	0.010	−0.006	0.022	0.005
	(0.793)	(−0.524)	(1.936)	(0.495)
Price				
Market capitalization	0.018	0.012	**0.033**	0.019
	(1.514)	(1.085)	**(2.854)**	(1.905)
Momentum	0.003	−0.001	0.003	−0.003
	(0.573)	(−0.260)	(0.590)	(−0.740)
Total volatility	−0.003	−0.005	−0.001	−0.008
	(−0.272)	(−0.541)	(−0.095)	(−0.948)
Industry-month FE	Yes	Yes	Yes	Yes
N	35,139	34,902	35,139	35,139
Adjusted R^2	0.059	0.057	0.047	0.052

Note: We measured disagreement as the standard deviation of all firm-level ratings available for a given firm at a given time. The explanatory variables are the following: tangibility, current ratio, leverage, gross profitability (see Novy-Marx 2013), the Herfindahl–Hirschman index (HHI) for measuring industry concentration, multisegment, no credit rating, institutional ownership, number of analysts, dispersion of analyst forecasts of the firm's one-year-ahead earnings forecasts (StD. EPS; see Diether et al. 2002), book-to-market ratio (see Fama and French 1995), market capitalization (see Banz 1981), momentum (see Jegadeesh and Titman 1993), and total volatility (see Ang, Hodrick, Xing, and Zhang 2006). We included industry-month fixed effects (FE). The *t*-statistics are based on double-clustered standard errors (month and firm). Bold indicates significance at the 5% level.

emissions) and thus potentially more easily measurable environmental ratings. Moreover, higher levels of institutional ownership are also associated with higher disagreement in the environmental rating. Firms with high book-to-market ratios display higher disagreement in their social ratings.

ESG Rating Disagreement and Stock Returns. We now turn to our main research question and examine the relationship between stock returns and ESG rating disagreement. As in the previous analyses, we used pooled panel regressions with standard errors double clustered at the firm and month level. We used monthly stock returns as the dependent variable in the regressions. In addition to our main disagreement-related explanatory variables, which we measured as the standard deviations of ratings available for a given firm at monthly intervals and denote as Disp, we included industry-month fixed effects. We also controlled for standard characteristics that have been found to explain the cross-section of stock returns: market capitalization (Banz 1981), book-to-market ratio (Fama and French 1995), gross profitability (Novy-Marx 2013), momentum (Jegadeesh and Titman 1993), dispersion of analyst forecasts of the firm's one-year-ahead earnings (Diether et al. 2002), the firm's beta (Frazzini and Pedersen 2014), and total volatility (Ang et al. 2006). Conceptually, pooled panel regressions with industry-month fixed effects are similar to Fama and MacBeth (1973) type regressions with industry dummies. Because our sample period is relatively short, and also because ESG rating disagreement varies among sectors (see Figure 1), controlling for differences at the industry level was important.

The coefficient estimates for the main explanatory variable (Disp) and the control variables appear in **Table 4**. We also show *t*-statistics based on double-clustered standard errors.

Table 4 reports a positive and significant coefficient estimate for the total ESG rating disagreement proxy, suggesting that firms with higher disagreement tend to have higher stock returns. The regression in column 1 shows an estimated coefficient of about 0.7 for disagreement about the total ESG rating. In terms of economic magnitude, consider a firm moving from the first quartile (0.14) to the third quartile (0.25) of the ESG rating disagreement distribution. Such a move would imply an increase of about 92 bps in annual stock returns [(0.25 – 0.14) × 0.7 × 12], an estimate that seems plausible in terms of magnitude.

In column 2, we find a coefficient estimate for disagreement about the environmental rating of about 1.0, which is both economically larger and more significant (*t*-statistic = 2.38) than the estimate for disagreement about the total rating. The coefficient for disagreement about the social rating (see column 3) is positive (0.6) but not significant at conventional levels (*t*-statistic = 1.515). The coefficient estimate for disagreement about the governance rating (see column 4) is small (–0.06) and insignificant (*t*-statistic = –0.136). We conclude that disagreement about the environmental rating primarily drives our results, with an interquartile range increase in disagreement about the environmental rating resulting in an increase of about 132 bps [(0.25 – 0.14) × 1 × 12] in a firm's annual cost of equity capital.

Given that ratings from Inrate and FTSE were not available for the entire sample period, we re-estimated the regressions by using a disagreement measure based only on Asset4, Sustainalytics, Bloomberg, MSCI KLD, and MSCI IVA. The results are reported in **Table B1** of Appendix B and continue to show a strong positive relationship between ESG rating disagreement and stock returns. In fact, the results in these regressions are stronger both statistically and economically speaking than the results shown in Table 4 for disagreement about the E and S pillars separately. The *t*-statistic for disagreement about the environmental rating increases to 3.01, and disagreement about the social rating has a *t*-statistic of 1.72.

The reader may wonder why the control variables in the return regressions of Table 4 do not turn out to be significant. This result might have a variety of reasons. First, the control variables are known return predictors, and there is evidence of lower post-publication return predictability for them (McLean and Pontiff 2016), especially in the United States (Jacobs and Müller 2020). Other research also has shown that return predictability fell sharply after 2003 (Green, Hand, and Zhang 2017). In a similar spirit, Chordia, Subrahmanyam, and Tong (2014) showed that capital market anomalies have attenuated in recent periods, coinciding with periods that have been accompanied by significant liquidity increases. Given the combined findings of these studies, and noting that we examined a sample of highly liquid S&P 500 firms since 2010, insignificant control variables are perhaps not that surprising.

Overall, we conclude that higher stock returns for firms with higher ESG rating disagreement is consistent with the view that risk-averse investors perceive

Table 4. Stock Returns and ESG Rating Disagreement, January 2010–December 2017 (*t*-statistics in parentheses)

	Dependent Variable: Returns			
Variable	Total (1)	Environmental (2)	Social (3)	Governance (4)
Disp	**0.698**	**1.012**	0.630	−0.059
	(1.995)	**(2.375)**	(1.515)	(−0.136)
Market capitalization	−0.134	−0.128	−0.135	−0.119
	(−0.726)	(−0.686)	(−0.725)	(−0.635)
Book-to-market ratio	0.216	0.233	0.218	0.226
	(0.973)	(1.048)	(0.984)	(1.021)
Gross profitability	0.231	0.246	0.234	0.222
	(0.772)	(0.825)	(0.781)	(0.745)
Momentum	0.360	0.348	0.359	0.361
	(1.114)	(1.064)	(1.109)	(1.120)
StD. EPS	−0.232	−0.193	−0.233	−0.228
	(−1.337)	(−1.116)	(−1.344)	(−1.313)
Beta	0.165	0.188	0.168	0.158
	(0.330)	(0.376)	(0.336)	(0.316)
Total volatility	−0.081	−0.115	−0.084	−0.073
	(−0.227)	(−0.328)	(−0.237)	(−0.206)
Industry-month FE	Yes	Yes	Yes	Yes
N	42,058	41,786	42,058	42,058
Adjusted R^2	0.347	0.348	0.347	0.347

Note: We measured ESG rating disagreement by the standard deviation of ratings available for a given firm at a given point in time, Disp. Included also are industry-month fixed effects and controls for standard characteristics that have been found to explain stock returns. The *t*-statistics are based on double-clustered standard errors (month and firm). Bold indicates significance at the 5% level.

a dispersed ESG performance of a given firm as an additional source of risk (or uncertainty) that commands a separate risk premium.

Portfolio Sorts on ESG Rating Disagreement. In the previous subsection, we documented a positive relationship between ESG rating disagreement and stock returns. To evaluate whether ESG rating disagreement provides a profitable signal for investing, we now report our implementation of portfolios sorted on the basis of ESG rating disagreement.

In the portfolio sorts, we used industry-adjusted ESG rating disagreement as the sorting variable. We did so primarily to rule out the possibility that our results would be driven by industry effects. As Figure 1 shows, average ESG rating correlations, and thus disagreement, exhibit important variation by industry, and we wanted to ensure that the portfolio compositions in the sorts were not biased by these industry differences in ESG rating disagreement. Note that in the return regressions of Table 4, we also controlled for industries by including industry-month fixed effects, implying that our insights and results are driven by differences in disagreement *within*, and not *between*, industries.

Specifically, for the portfolio sorts, we calculated the industry-adjusted ESG rating disagreement for a given firm in a given month by simply demeaning stock-level ESG rating disagreement using

the average ESG rating disagreement in the firm's Fama–French 12-industry classification in a given month. We denote this variable as Disp_adj, and we sorted all stocks in the sample into five quintile groups based on their industry-adjusted ESG rating disagreement. **Table 5** shows the results.

For each of the four rating dimensions, we report in Table 5 results for the equally weighted portfolios of stocks (quintile) with the lowest industry-adjusted rating disagreement (Low Disp_adj), the highest industry-adjusted rating disagreement (High Disp_adj), and the long–short portfolio of high- and low-disagreement stocks. Returns are all expressed in excess of the risk-free rate. We report mean monthly returns, the median number of firms for the low- and high-disagreement portfolios, standard deviations, Sharpe ratios, and alphas computed from four different factor models. We observe in Panel A of Table 5 that for the total rating, the equally weighted long–short portfolio of high- and low-disagreement stocks generated a mean raw monthly return of 21 bps (with a *t*-statistic of 2.2). The factor model alphas are of a similar magnitude, with average monthly alphas ranging between 23 bps and 27 bps (and *t*-statistics between 2.0 and 2.7). Furthermore, we observe that a long-only strategy of high-disagreement stocks generated a raw monthly mean return of 134 bps (with a *t*-statistic of 3.5). The alphas of the long-only strategy, ranging between 14 bps and 21 bps (with *t*-statistics between 1.6 and 2.4), are somewhat lower.

For the environmental dimension (Panel B of Table 5), we observe similar results. The long–short portfolio generated a mean monthly raw return of 21 bps (with a *t*-statistic of 2.0). The factor model alphas range between 21 bps and 25 bps (with *t*-statistics between 2.0 and 2.5). The long-only strategy also shows a high average monthly raw return, about 140 bps (*t*-statistic of 3.5), and also high average monthly factor alphas, ranging from 16 bps to 21 bps (with *t*-statistics between 1.8 and 2.3). For disagreement about the social and governance ratings, we did not obtain any significant raw returns or alphas from the long–short strategies. Taken together, the insights from the portfolio sorts in Table 5 are consistent with the evidence presented in the regression analysis shown in Table 4 and emphasize the prominent role played by environmental rating disagreement in explaining stock returns.

An important question is whether these strategies are implementable in a way that survives accounting for trading costs. To gain a realistic impression as to whether they are, we followed the approach advocated by Novy-Marx and Velikov (2016). These authors used the effective bid–ask spread measure of Hasbrouck (2009) to proxy for trading costs. Hasbrouck suggested a Bayesian Gibbs sampler approach to the Roll (1984) model of price dynamics. Hasbrouck showed that his estimate of the average effective cost is comparable to estimates from high-frequency trade and quote (TAQ) data and reported a 0.965 Pearson correlation of his measure with the TAQ value. Novy-Marx and Velikov noted that this measure does not account for the price impact of large trades, but they nevertheless considered it an appropriate trading cost measure (for details, see Novy-Marx and Velikov, p. 108).

To calculate the actual trading costs for each portfolio, we subtracted trading costs from portfolio returns each time a given stock was introduced into a portfolio or withdrawn from it.

Table 6 displays the results of the long–short strategies after accounting for transaction costs. Note that the raw returns, Sharpe ratios, and alphas are somehow diminished by trading costs but still sizable. For example, for the total rating disagreement, the monthly raw return of the long–short portfolio of high- and low-disagreement stocks drops from 21 bps in Table 5 to 18 bps in Table 6. The *t*-statistic drops from 2.2 to 1.9. The monthly alphas range from 19.5 bps to 23 bps (with *t*-statistics of 1.8 to 2.4). For environmental rating disagreement, we observe significant monthly alphas for the CAPM and the Fama–French three-factor and Carhart four-factor models. These alphas range between 17 bps and 21 bps (with *t*-statistics between 2.0 and 2.2). The long–short raw returns and the alphas from the Fama–French five factor model, however, have *t*-statistics of about 1.6 in both cases (and mean returns of about 17 bps). As in the case without transaction costs, the long–short portfolios for the social and the governance rating disagreements display insignificant raw returns and risk-adjusted alphas.

Possible Theoretical Explanations. A potential way of rationalizing the positive relationship between ESG ratings disagreement and stock returns may be found in the literature on heterogeneous beliefs in financial markets. Theoretical models of heterogeneous beliefs (e.g., Atmaz and Basak 2018) provide some predictions for empirical studies. Note, however, that these theories are based on beliefs about factors that affect the returns of firms (e.g., consensus EPS forecasts), not factors that may or may not have risk and return implications, such as

Table 5. Portfolio Sorts on Industry-Adjusted ESG Rating Disagreement, January 2010–December 2017 (*t*-statistics in parentheses)

	Return (1)	*N* (2)	StD. (3)	Sharpe Ratio (4)	CAPM (5)	FF3 (6)	Car (7)	FF5 (8)
A. Total								
Low Disp_adj	**1.124**	94	3.809	0.295	−0.088	−0.068	−0.055	−0.074
	(2.891)				(−1.260)	(−0.970)	(−0.794)	(−1.049)
High Disp_adj	**1.336**	94	3.769	0.355	0.144	0.164	**0.211**	0.165
	(3.474)				(1.561)	(1.792)	**(2.425)**	(1.611)
H–L Disp_adj	**0.212**		0.942	0.225	**0.232**	**0.232**	**0.267**	**0.239**
	(2.209)				**(2.151)**	**(2.192)**	**(2.664)**	**(2.044)**
B. Environmental								
Low Disp_adj	**1.186**	93	4.057	0.292	−0.091	−0.043	−0.016	−0.036
	(2.866)				(−0.902)	(−0.500)	(−0.187)	(−0.455)
High Disp_adj	**1.397**	93	3.906	0.358	0.163	**0.183**	**0.212**	**0.178**
	(3.504)				(1.848)	**(2.040)**	**(2.274)**	**(1.982)**
H–L Disp_adj	**0.211**		1.026	0.205	**0.254**	**0.226**	**0.228**	**0.213**
	(2.010)				**(2.482)**	**(2.226)**	**(2.322)**	**(2.005)**
C. Social								
Low Disp_adj	**1.283**	93	3.841	0.334	0.071	0.121	0.139	0.116
	(3.273)				(0.865)	(1.734)	(1.943)	(1.950)
High Disp_adj	**1.330**	94	3.996	0.333	0.067	0.096	0.134	0.081
	(3.262)				(0.658)	(1.077)	(1.748)	(0.998)
H–L Disp_adj	0.047		0.965	0.049	−0.004	−0.024	−0.004	−0.035
	(0.480)				(−0.058)	(−0.338)	(−0.073)	(−0.479)
D. Governance								
Low Disp_adj	**1.246**	94	3.888	0.320	0.013	0.050	0.071	0.057
	(3.140)				(0.129)	(0.535)	(0.777)	(0.702)
High Disp_adj	**1.284**	94	3.794	0.338	0.088	0.118	**0.167**	**0.120**
	(3.315)				(1.320)	(1.871)	**(2.749)**	**(2.577)**
H–L Disp_adj	0.038		1.006	0.037	0.075	0.069	0.096	0.063
	(0.366)				(0.742)	(0.680)	(0.954)	(0.696)

Note: Reported are mean returns (Return), the median number of observations for the high- and low-disagreement portfolios (*N*), the standard deviations of returns (StD.), and the Sharpe ratio. In addition, we report alphas of the capital asset pricing model (CAPM) and the Fama–French three-factor, Carhart four-factor, and Fama–French five-factor models. The sample includes 96 monthly time-series observations. Portfolios were formed each January with disagreement values from December of the preceding year. Returns are reported in percentages. The *t*-statistics for the factor model alphas are based on Newey–West (1987) standard errors with 12 lags. Bold indicates significance at the 5% level.

ESG ratings. Nonetheless, we think that discussing our findings in the light of this literature will be useful in our setting without attempting to test a specific theory.

In an important paper, Atmaz and Basak (2018, p. 1241) argued that ". . . dispersion represents an additional risk for investors and therefore investors demand a higher return to hold the stock when

Table 6. Portfolio Sorts on Industry-Adjusted ESG Rating Disagreement Adjusted for Trading Costs, January 2010–December 2017 (*t*-statistics in parentheses)

	Return (1)	*N* (2)	StD. (3)	Sharpe Ratio (4)	CAPM (5)	FF3 (6)	Car (7)	FF5 (8)
A. Total								
Low Disp_adj	**1.119**	93	3.797	0.295	−0.089	−0.068	−0.056	−0.078
	(2.887)				(−1.288)	(−0.988)	(−0.820)	(−1.110)
High Disp_adj	**1.330**	94	3.762	0.354	0.139	0.159	**0.206**	0.162
	(3.465)				(1.500)	(1.715)	**(2.365)**	(1.567)
H–L Disp_adj	0.181		0.934	0.194	0.196	0.195	**0.227**	0.203
	(1.897)				(1.890)	(1.906)	**(2.370)**	(1.830)
B. Environmental								
Low Disp_adj	**1.178**	93	4.054	0.291	−0.097	−0.050	−0.023	−0.041
	(2.848)				(−0.977)	(−0.575)	(−0.276)	(−0.510)
High Disp_adj	**1.379**	93	3.889	0.355	0.150	**0.169**	**0.197**	0.162
	(3.474)				(1.775)	**(1.964)**	**(2.247)**	(1.882)
H–L Disp_adj	0.168		1.025	0.164	**0.214**	0.185	0.185	0.166
	(1.609)				**(2.155)**	(1.861)	(1.939)	(1.594)
C. Social								
Low Disp_adj	**1.270**	93	3.835	0.331	0.059	0.109	0.126	0.102
	(3.243)				(0.742)	(1.589)	(1.758)	(1.690)
High Disp_adj	**1.320**	93.5	3.966	0.333	0.065	0.089	0.125	0.075
	(3.260)				(0.681)	(1.052)	(1.693)	(0.985)
H–L Disp_adj	0.017		0.958	0.018	−0.029	−0.055	−0.038	−0.066
	(0.172)				(−0.443)	(−0.777)	(−0.630)	(−0.954)
D. Governance								
Low Disp_adj	**1.225**	93	3.929	0.312	−0.021	0.020	0.040	0.025
	(3.054)				(−0.201)	(0.203)	(0.421)	(0.297)
High Disp_adj	**1.266**	93	3.821	0.331	0.062	0.091	**0.141**	0.090
	(3.247)				(0.934)	(1.444)	**(2.274)**	(1.916)
H–L Disp_adj	0.005		1.031	0.005	0.045	0.033	0.060	0.022
	(0.044)				(0.401)	(0.302)	(0.562)	(0.227)

Note: See the notes to Table 5. We adjusted returns for trading costs according to Novy-Marx and Velikov (2016). Bold indicates significance at the 5% level.

dispersion is higher." In their setting, this impact may have been, however, attenuated (reinforced) in the presence of investors' excessive optimism (pessimism). Along the same lines, Anderson et al. (2005) found that adding empirical factors based on analysts' EPS forecast dispersion enhanced explanatory power in explaining S&P 500 excess returns after accounting for standard market risk factors. Extended to our context, that approach would imply that ESG rating disagreement risk is priced

in expected stock returns in addition to the market risk exposure of stocks. In other words, our finding of a positive relationship between stock returns and ESG rating disagreement is in line with the risk-based explanation for analysts' EPS disagreement empirically documented by Anderson et al. According to this risk-based explanation, more total (or environmental) rating disagreement implies more uncertainty about the ESG performance of a given firm and thus would be perceived as a separate source of risk that commands a risk premium if investors are risk averse. We found supportive empirical evidence that this explanation is correct for environmental and total rating disagreement. So, we have thereby added to this stream of literature by empirically documenting the effects of heterogeneity in beliefs about nonfinancial information on stock excess returns.

Another possible explanation for our findings is that disagreement about a firm's ESG rating is a proxy for ESG uncertainty and therefore captures a specific form of (Knightian) uncertainty. Whereas risk is associated with the uncertain outcome of a known probability distribution of returns, ambiguity (or Knightian uncertainty) is associated with uncertainty regarding the probability distribution itself (see, for example, Viale et al. 2014). Theoretical papers have advocated using a two-factor model, with one factor a proxy for risk and the other factor a proxy for uncertainty (see, in particular, Kogan and Wang 2003). Anderson, Ghysels, and Juergens (2009) estimated a two-factor model with a factor proxy for uncertainty based on professional forecasters' disagreement. Viale et al. further showed that uncertainty (or ambiguity) is priced in the cross-section of stock returns and not subsumed by standard risk factors. Hence, in the context of our setting, a possible explanation for the positive relationship between ESG rating disagreement and stock returns is that our ESG rating disagreement measure is a proxy for uncertainty regarding ESG information—in particular, about environmental information.

ESG Rating Disagreement and Equity Risk. For completeness, we examine in this subsection whether ESG rating disagreement affects standard equity risk measures. An examination of the relationship between equity risk and ESG disagreement is natural because prior research highlights a robustly negative relationship between stock-level risk and the quality of a firm's ESG rating (Dunn, Fitzgibbons, and Pomorski 2018; Hoepner, Oikonomou, Sautner, Starks, and Zhou 2019; Gibson et al. 2020). Hence, we regressed standard risk measures on our ESG rating disagreement proxies and a set of common control variables.[15] **Table 7** reflects our focus on the three risk measures—*total volatility*, *idiosyncratic volatility*, and *beta*. Following prior studies, we also analyzed whether downside risk, as measured by the lower partial moment (LPM), is related to the level of ESG rating disagreement (see Hoepner et al. for details on the relationship between ESG policies and LPMs).

Table 7 shows that some risk measures are positively related to ESG rating disagreement—in particular, to disagreement about the social dimension. For example, total volatility, idiosyncratic volatility, and LPM all appear to be positively related to disagreement about the social rating (see column 3). In addition, we found some evidence that idiosyncratic and downside risk are related to total rating disagreement.

We believe that in light of the second theoretical explanation provided in the previous subsection ("Possible Theoretical Explanations"), to observe that ESG rating disagreement is only weakly related to standard risk measures is not surprising. The uncertainty-based explanation is consistent with the fact that ESG rating disagreement has little power in explaining standard risk measures. Therefore, our results suggest that ESG information uncertainty offers explanatory power beyond traditional risk factors.

Study Limitations. Our study has several limitations. First, the fact that we focused on the S&P 500 universe leaves open the question of whether our results can be transposed to other stock markets located outside of the United States. Also unclear is whether and how the conclusions hold for larger cross-sections of stocks. Second, to encompass as many data providers as possible, we had to work with a limited time period. Thus, the power of our tests might be an issue. This issue may explain why we found no significant effects for disagreement about the S and G ratings. Third, some of the data vendors may have changed their rating methods during our sample period, which would create additional biases when measuring ESG rating disagreement. Finally, we focused only on stocks, but exploring whether our results can be transposed to fixed-income securities that have ESG ratings might be interesting.

Table 7. Risk and ESG Rating Disagreement, January 2010–December 2017 (*t*-statistics in parentheses)

Pillars:	Total (1)	Environmental (2)	Social (3)	Governance (4)
Dependent variable: Total volatility				
Disp	0.002	−0.001	**0.003**	−0.001
	(1.614)	(−0.667)	**(2.124)**	(−0.949)
Controls	Yes	Yes	Yes	Yes
N	40,519	40,365	40,519	40,519
Adjusted R^2	0.573	0.571	0.573	0.573
Dependent variable: Idiosyncratic volatility				
Disp	0.002	−0.001	**0.003**	−0.001
	(1.710)	(−0.463)	**(2.223)**	(−0.819)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.477	0.474	0.477	0.476
Dependent variable: Beta				
Disp	0.075	−0.159	0.075	−0.088
	(0.690)	(−1.469)	(0.777)	(−0.959)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.441	0.442	0.441	0.441
Dependent variable: LPM				
Disp	0.002	0.000	**0.002**	−0.001
	(1.915)	(−0.271)	**(1.989)**	(−0.865)
Controls	Yes	Yes	Yes	Yes
N	40,515	40,362	40,515	40,515
Adjusted R^2	0.519	0.517	0.519	0.518

Note: For the calculation of idiosyncratic volatility and stock market beta, we used the CAPM. The *t*-statistics are based on double-clustered standard errors (month and firm). Bold indicates significance at the 5% level.

Conclusion

Recently, the issue of ESG rating disagreement has received considerable attention from the financial press and from practitioner and policy-making circles. In addition, ESG rating disagreement has important implications for the generalization of academic research findings and is creating challenges for asset managers in their efforts to implement ESG investment strategies. We have provided a first step toward a better understanding of the real consequences of ESG rating disagreement by studying the impact on stock returns. Specifically, we found that stock returns are positively related to ESG rating disagreement—in particular, environmental rating disagreement. We did not assume that the environmental pillar is, per se, more important than the social and the governance pillars but found its prominence as a result of the analysis. In addition, academics (see Bolton and Kacperczyk forthcoming) and practitioners (see the survey by BlackRock 2020) recently emphasized that environmental risks are important, which may point to the fact that disagreement about the environmental rating dimension is the only one priced so far by investors with ESG preferences.

Our results have important practical consequences. First, our analysis shows that financial analysts who value the equity of firms should incorporate the effects of ESG rating disagreement and adjust estimates of firms' equity cost of capital upward. Second, CFOs deciding about the allocation of capital expenditures should consider ESG rating disagreement in their capital budgeting decisions because such disagreement raises the investment threshold for firms subject to substantial total (and environmental) rating disagreement. Third, our evidence that ESG rating disagreement varies across industries is an important insight for financial analysts, who often focus on specific industries.

Finally, our analysis also has important implications for asset owners and investment managers who implement responsible investment strategies. In today's responsible investment landscape, two strategies are currently very popular—screening and ESG integration. If asset managers and investment managers following these strategies wish to optimize financial performance while investing responsibly, they should care about ESG rating disagreement and its impact on stock returns. Indeed, our results suggest that with positive (negative) screening, managers should buy (sell) primarily those stocks that, for a given high (low) ESG rating, command the lowest (highest) level of ESG disagreement. This practice should allow positive (negative) screeners to mitigate the adverse impact of ESG rating disagreement on the expected future returns of their buy (sell) orders. Similarly, an ESG integration strategy may fail to deliver its financial promises if it does not search simultaneously for those stocks that have superior ESG ratings *and* embed the lowest level of ESG rating disagreement within an industry. Indeed, controlling for a low level of ESG rating disagreement will allow investors who integrate ESG criteria in their stock selection processes to avoid a subsequent unintended stock price decline.

Appendix A. Sample Selection, Financial Data, Dataset Matching, and Variable Definitions

Sample Selection

To test our hypotheses, we constructed a representative and homogeneous sample over the longest possible period. We faced the challenge that the availability of ESG data is restricted in both the cross-section and the time series. In other words, ESG data are often available only for the largest firms and for recent years. To use a sample as homogeneous as possible and to maximize the number of available ESG ratings per firm, we restricted ourselves to firms belonging to the S&P 500 and considered a sample period of eight years, from 2010 to 2017. See Table A1 for an overview of the variables used in this study.

Financial Data

We used financial data from CRSP and accounting data from Standard & Poor's Compustat. For each stock, we calculated idiosyncratic volatility, total volatility, and the stock market beta at the end of each month for up to 250 daily observations (we required a minimum of 60 daily observations). We calculated market capitalization as (adjusted) total shares outstanding times stock price, both at the end of the month. We calculated the momentum signal at time t as the continuously compounded returns from month $t - 2$ to month $t - 12$. Book value of equity is the sum of shareholders' equity, deferred taxes, and the investment tax credit minus preferred stock.[16] Only firms with a positive book value were selected for the sample. Following Novy-Marx (2013), we calculated gross profitability as total revenues minus cost of goods sold, divided by total assets. In addition, we matched the dispersion in EPS forecasts for one-year-ahead earnings from IBES (Diether et al. 2002).

Dataset Matching

A challenge for constructing a dataset from many subsets of data is to properly match the different sets. We matched on the basis of three identifiers: (1) CUSIP, (2) ISIN, and (3) company name. The CUSIP code was available for all the providers except Inrate.[17] Because the ISIN code was available for Inrate, however, we extracted the CUSIP code from the ISIN code. Note that we used only the first six CUSIP characters for matching (known as the *issuer* identifier). Characters seven and eight identify the specific issue (for example, if the seventh character is 1 and the eighth character is 0, this indicates common equity), and the ninth character is a check digit. The ISIN code was available for all providers except MSCI KLD. For the CRSP/Compustat data, we retrieved the ISIN number from the CUSIP code

and the current ISO (International Organization for Standardization) country code of incorporation (fic).[18] To perform the merge with the company names, we first converted the original names of the providers by using some common abbreviations to avoid rather trivial mismatches. We used the unique union of all three matching procedures to compile our sample.

To construct the sample, we also required that at least three rating observations be available for each company. This choice provided us with an internally consistent sample, and it was not overly restrictive.

In addition, we used a monthly frequency for our sample. Asset4, Sustainalytics, FTSE, and MSCI IVA already provided data at a monthly frequency; Inrate provided rating updates on a semiannual basis for the years 2015 and 2016; and Bloomberg and MSCI KLD provided data on a yearly frequency. To convert from a semiannual or annual frequency, we simply used the respective annual or semiannual value for the whole period. Note that most ratings (also for the providers with a monthly frequency) change infrequently; most ratings are constant for about one year, but some are constant for longer periods.[19]

Table A1. Variables Overview

Variables	Description	Details	Source
ESG rating disagreement variables			
Disp	Standard deviation of all firm-level ratings	To compute standard deviations, we adjusted the raw ratings as follows: We calculated the percentile ranks and used these as adjusted scores.	Thomson Reuters/ Refinitiv, Sustainalytics, Inrate, Bloomberg, FTSE MSCI
Disp_adj	Industry-adjusted standard deviation of firm-level ratings	A firm's ESG rating disagreement adjusted by the average rating disagreement in the firm's Fama–French 12 industry classification.	Thomson Reuters/ Refinitiv, Sustainalytics, Inrate, Bloomberg, FTSE MSCI, Fama and French data library
Additional independent variable(s)			
Return	Stock returns	Monthly stock returns.	CRSP
Control variables			
Tangibility	Tangibility	Property, plant, and equipment divided by total assets.	Compustat
Current ratio	Current ratio	Current assets divided by current liabilities.	Compustat
Leverage	Leverage	Long-term debt plus debt in current liabilities divided by total assets.	Compustat
Gross profitability	Gross profitability	Revenues minus costs of goods sold divided by total assets.	Compustat
HHI	Herfindahl–Hirschman Index (HHI) based on book equity	The HHI measures industry concentration by using book equity and the 2-digit SIC level.	Compustat
Multisegment	Multisegment	Dummy variable that is 1 if the firm operates in more than one segment.	Compustat Segments Data
No credit rating	Missing credit rating	Dummy variable that is 1 if no credit rating is available.	Compustat Company S&P Credit Ratings
Institutional ownership	Institutional ownership	Percentage of institutional ownership.	Thomson Reuters Institutional (13f) Holdings

(continued)

Table A1. Variables Overview (continued)

Variables	Description	Details	Source
Number of analysts	Number of analysts	Number of analysts, based on IBES summary files.	IBES
StD. EPS	Dispersion of analyst EPS forecasts	Dispersion of analyst forecasts of the firm's one-year-ahead earnings forecasts, measured by standard deviation.	IBES
Book-to-market ratio	Book-to-market ratio	Book equity (shareholders' equity plus deferred taxes plus investment tax credit minus preferred stock [redemption value, liquidation value, or carrying value], based on availability) divided by market capitalization.	Compustat, CRSP
Market capitalization	Market capitalization	Absolute value of stock price multiplied by shares outstanding.	CRSP
Momentum	Momentum	Cumulative returns of the most recent 12 months, excluding the most recent month for each firm (from month t – 12 to month t – 2).	CRSP
Risk measures			
Total volatility	Total volatility	Standard deviation computed from the most recent 250 daily return observations.	CRSP
Idiosyncratic volatility	Idiosyncratic volatility	Total volatility minus systematic volatility (explained by the CAPM) from the most recent 250 daily return observations.	CRSP
Beta	Firm's beta	Market beta computed from the most recent 250 daily return observations.	CRSP
Downside risk measures			
LPM	Lower partial moment (log-transformed)	The lower partial moment is the square root of the standard deviation of the negative return part of the distribution. For details, see Hoepner et al. (2019).	CRSP

Note: We classified the variables into five groups: ESG rating disagreement variables, additional independent variable(s), control variables, risk measures, and downside risk measures.

Table A2. Summary Statistics

Variable	*N* (1)	Mean (2)	StD. (3)	Min (4)	Max (5)	Median (6)	Skew (7)	Kurt (8)
ESG rating disagreement variables								
Disp (T)	51,206	0.200	0.085	0.000	0.560	0.194	0.386	–0.055
Disp (E)	50,904	0.204	0.081	0.000	0.547	0.202	0.287	0.075
Disp (S)	50,877	0.222	0.083	0.000	0.544	0.220	0.131	–0.285
Disp (G)	50,877	0.241	0.081	0.000	0.535	0.242	–0.007	–0.315

(continued)

Table A2. Summary Statistics (continued)

Variable	*N* (1)	Mean (2)	StD. (3)	Min (4)	Max (5)	Median (6)	Skew (7)	Kurt (8)
Other variables								
Return	51,152	0.012	0.080	−0.748	1.274	0.013	0.388	6.796
Market capitalization	51,178	0.030	0.052	0.000	0.882	0.013	4.992	38.167
Book-to-market ratio	49,783	0.762	2.381	0.001	108.763	0.444	20.718	621.428
Gross profitability	49,783	0.287	0.219	−1.143	1.405	0.253	1.035	2.459
Momentum	47,736	0.150	0.305	−0.925	4.870	0.139	1.320	10.567
StD. EPS	44,310	0.822	17.904	0.000	961.040	0.060	31.516	1,147.771
Beta	51,094	1.075	0.394	−1.738	3.787	1.039	0.712	1.461
Total volatility	51,178	0.018	0.009	0.001	0.226	0.016	4.899	65.671
Tangibility	49,304	0.255	0.244	0.000	0.946	0.160	1.022	−0.102
Current ratio	42,716	1.832	1.188	0.205	11.966	1.490	2.673	11.596
Leverage	49,585	0.246	0.157	0.000	0.960	0.238	0.591	0.561
HHI	49,783	0.074	0.078	0.009	0.744	0.044	2.810	11.632
Multisegment	18,212	0.562	0.496	0.000	1.000	1.000	−0.249	−1.938
No credit rating	51,206	0.199	0.399	0.000	1.000	0.000	1.511	0.284
Institutional ownership	50,811	0.751	0.221	0.000	1.748	0.801	−1.811	3.800
Number of analysts	48,920	0.189	0.078	0.010	0.560	0.180	0.445	0.481
Idiosyncratic volatility	51,094	0.014	0.008	0.001	0.226	0.012	6.994	135.876
LPM	51,094	0.012	0.006	0.001	0.103	0.011	3.054	21.447

Note: We report the number of observations (*N*) and the sample means and standard deviations, as well as minimums, maximums, medians, skewness, and kurtosis. Table A1 provides a detailed description of the variables.

Appendix B. Robustness Check: Sample without FTSE and Inrate

Table B1. Stock Returns and ESG Rating Disagreement for Sample without FTSE and Inrate (*t*-statistics in parentheses)

	Dependent Variable: Returns			
Variable	Total (1)	Environmental (2)	Social (3)	Governance (4)
Disp	0.626	**1.217**	0.698	−0.376
	(1.762)	**(3.005)**	(1.720)	(−1.023)
Market capitalization	−0.131	−0.108	−0.129	−0.111
	(−0.710)	(−0.579)	(−0.694)	(−0.592)

(continued)

Table B1. Stock Returns and ESG Rating Disagreement for Sample without FTSE and Inrate (*t*-statistics in parentheses) (continued)

	Dependent Variable: Returns			
Variable	Total (1)	Environmental (2)	Social (3)	Governance (4)
Book-to-market ratio	0.218	0.240	0.228	0.229
	(0.983)	(1.076)	(1.026)	(1.036)
Gross profitability	0.225	0.256	0.240	0.219
	(0.751)	(0.860)	(0.799)	(0.735)
Momentum	0.361	0.348	0.357	0.361
	(1.114)	(1.063)	(1.102)	(1.120)
StD. EPS	−0.230	−0.194	−0.234	−0.225
	(−1.325)	(−1.119)	(−1.355)	(−1.298)
Beta	0.158	0.202	0.158	0.155
	(0.316)	(0.403)	(0.317)	(0.310)
Total volatility	−0.078	−0.101	−0.082	−0.073
	(−0.220)	(−0.288)	(−0.231)	(−0.206)
Industry-month FE	Yes	Yes	Yes	Yes
N	42,032	41,738	42,032	42,032
Adjusted R^2	0.347	0.348	0.347	0.347

Note: See the notes to Table 4. Bold indicates significance at the 5% level.

Editor's Note

Submitted 6 February 2021

Accepted 29 July 2021 by William N. Goetzmann

Notes

1. Valuation mistake = [Equity(no disagreement) − Equity(disagreement)]/Equity(disagreement) − 1 = [FCFE/(r_E − g)]/[FCFE/(r_E + 0.0092 − g)] = [100/(0.05)]/[100/(0.05 + 0.0092)] − 1 = 0.184 (or 18.4%).

2. See, for instance, Hong and Kostovetsky (2012); Krüger (2015); Lins, Servaes, and Tamayo (2017); Liang and Renneboog (2017); Gibson, Krueger, and Mitali (2020); Dyck, Lins, Roth, and Wagner (2019).

3. See, for instance, Chatterji, Levine, and Toffel (2009); Bouten, Cho, Michelon, and Roberts (2017); Delmas, Etzion, and Nairn-Birch (2013).

4. The concept of a *common theorization* refers to the idea that raters (or information intermediaries) agree on a common definition of CSR. Absence of *commensurability* captures the idea that different raters do not use the same measures when quantifying the same feature.

5. In a recent study, Lopez-de-Silanes, McCahery, and Pudschedl (2019) found evidence that firms with good ESG scores may disclose more information.

6. Asset4 was acquired by Thomson Reuters in 2009, but the ESG data were made available under the old name of Asset4. After the acquisition, the name changed to Thomson Reuters ESG Scores. Because the name Asset4 is widely known, however, we used the old name for simplicity. Note that as of 2018, the ESG ratings data of Thomson Reuters are part of Refinitiv and now also known as Refinitiv ESG.

7. After acquiring about a 40% stake in Sustainalytics in 2017, Morningstar purchased the remaining approximately 60% of Sustainalytics equity in 2020 (see https://bit.ly/3oXCgxM).

8. The data from KLD originate from Kinder, Lydenberg, Domini (KLD) & Co., Inc., which was acquired by RiskMetrics in 2009. In 2010, MSCI acquired RiskMetrics. Eccles, Lee, and Stroehle (2019) have provided details on the history of KLD. We refer to these data as either KLD or MSCI KLD.

9. The MSCI IVA dataset was initially created by Innovest Strategic Value Advisors, which was also acquired by RiskMetrics in 2009 before RiskMetrics was taken over by MSCI (see Eccles et al. 2019 for details).

10. Each provider, with the exception of Bloomberg, had a rather constant number of observations for the different scores they issue. Bloomberg had substantially lower coverage for environmental ratings.

11. Available at http://www.ghgprotocol.org.

12. In addition, evidence from both academics and practitioners indicates that environmental risk matters to investors. Bolton and Kacperczyk (forthcoming) found a sizable risk premium for firms with high carbon emissions, suggesting that investors do care about carbon risk. In a recent survey by BlackRock (2020), 88% of the respondents placed climate risk at the top of their portfolio concerns.

13. The 12-industry classification may be found at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library/det_12_ind_port.html.

14. The reader might wonder why any S&P 500 firms would not have a credit rating. In general, firms without a credit rating do not seem to be exceptional. For example, in a sample of 12,312 firms, Avramov, Chordia, Jostova, and Philipov (2009) reported that 9,051 firms did not have a credit rating. In our sample, 194 out of a total of 553 firms did not have a credit rating for at least one month.

15. We included industry-month fixed effects and controlled for standard characteristics (not reported) that have been found to explain volatility (see, for example, Dennis and Strickland 2004)—namely, market capitalization (Banz 1981), leverage, business segment (a dummy variable was 1 for multisegment firms), percentage of institutional ownership, the ratio of mutual fund ownership to total institutional ownership, and turnover.

16. If available, we used the redemption value as preferred stock. Otherwise, we used the liquidating value or, if the liquidation value was also unavailable, the carrying value.

17. The MSCI KLD dataset seems to have some issues with the CUSIP code. The codes do not always have the same number of characters, and leading zeros are often truncated. Therefore, we filled in leading zeros if the number of characters was less than eight. Then, we added the self-computed check digit to the code if the eighth number was not the would-be check digit if there had been an additional leading zero (in that case, we added a leading zero) or the last two characters consisted of commonly used issue codes.

18. For US stocks, the ISIN number consists of the country code (first two characters), the CUSIP code (characters 3 to 11), and a check digit.

19. Because the providers change their ratings at different points in time, we argue that for our purposes, using a monthly frequency makes sense.

References

Anderson, E. W., E. Ghysels, and J. L. Juergens. 2005. "Do Heterogeneous Beliefs Matter for Asset Pricing?" *Review of Financial Studies* 18 (3): 875–924.

——. 2009. "The Impact of Risk and Uncertainty on Expected Returns." *Journal of Financial Economics* 94 (2): 233–63.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. "The Cross-Section of Volatility and Expected Returns." *Journal of Finance* 61 (1): 259–99.

Atmaz, A., and S. Basak. 2018. "Belief Dispersion in the Stock Market." *Journal of Finance* 73 (3): 1225–79.

Avramov, D., S. Cheng, A. Lioui, and A. Tarelli. 2021. "Sustainable Investing with ESG Rating Uncertainty." Working paper.

Avramov, D., T. Chordia, G. Jostova, and A. Philipov. 2009. "Dispersion in Analysts' Earnings Forecasts and Credit Rating." *Journal of Financial Economics* 91 (1): 83–101.

Banz, R. W. 1981. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics* 9 (1): 3–18.

Berg, F., J. F. Koelbel, and R. Rigobon. 2020 "Aggregate Confusion: The Divergence of ESG Ratings." Working paper.

BlackRock. 2020. "BlackRock Survey Shows Acceleration of Sustainable Investing." Press release. https://bit.ly/3upXdlZ.

Bolton, P., and A. Kacperczyk. Forthcoming. "Do Investors Care about Carbon Risk?" *Journal of Financial Economics*.

Bouten, L., C. H. Cho, G. Michelon, and R. W. Roberts. 2017. "CSR Performance Proxies in Large-Sample Studies: 'Umbrella Advocates', Construct Clarity and the 'Validity Police.'" Working paper.

Carhart, M. M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52 (1): 57–82.

Chatterji, A. K., R. Durand, D. I. Levine, and S. Touboul. 2016. "Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers." *Strategic Management Journal* 37 (8): 1597–614.

Chatterji, A. K., D. I. Levine, and M. W. Toffel. 2009. "How Well Do Social Ratings Actually Measure Corporate Social Responsibility?" *Journal of Economics & Management Strategy* 18 (1): 125–69.

Chordia, T., A. Subrahmanyam, and Q. Tong. 2014. "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" *Journal of Accounting and Economics* 58 (1): 41–58.

Christensen, D., G. Serafeim, and A. Sikochi. Forthcoming. "Why Is Corporate Virtue in the Eye of the Beholder? The Case of ESG Ratings." *Journal of Accounting Research*.

Delmas, M. A., D. Etzion, and N. Nairn-Birch. 2013. "Triangulating Environmental Performance: What Do Corporate Social Responsibility Ratings Really Capture?" *Academy of Management Perspectives* 27 (3): 255–67.

Dennis, P. J., and D. Strickland. 2004. "The Determinants of Idiosyncratic Volatility." Working paper.

Diether, K. B., C. J. Malloy, and A. Scherbina. 2002. "Differences of Opinion and the Cross Section of Stock Returns." *Journal of Finance* 57 (5): 2113–41.

Doyle, T. M. 2018. "Ratings That Don't Rate: The Subjective View of ESG Rating Agencies." Technical report. American Council for Capital Formation. https://bit.ly/2LBwvky.

Dunn, J., S. Fitzgibbons, and L. Pomorski. 2018. "Assessing Risk through Environmental, Social and Governance Exposures." *Journal of Investment Management* 16 (1): 4–17.

Dyck, A., K. V. Lins, L. Roth, and H. F. Wagner. 2019. "Do Institutional Investors Drive Corporate Social Responsibility? International Evidence." *Journal of Financial Economics* 131 (3): 693–714.

Eccles, R. G., L.-E. Lee, and J. C. Stroehle. 2019. "The Social Origins of ESG? An Analysis of Innovest and KLD." Working paper.

Fama, E. F., and K. R. French. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Finance* 50 (1): 131–55.

——. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1–22.

Fama, E. F., and J. D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy* 81 (3): 607–36.

Frazzini, A., and L. H. Pedersen. 2014. "Betting against Beta." *Journal of Financial Economics* 111 (1): 1–25.

Friede, G., T. Busch, and A. Bassen. 2015. "ESG and Financial Performance: Aggregated Evidence from More Than 2000 Empirical Studies." *Journal of Sustainable Finance & Investment* 5 (4): 210–33.

Gibson, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen. 2021. "Do Responsible Investors Invest Responsibly?" University of Geneva and University of Virginia.

Gibson, R., P. Krueger, and S. Mitali. 2020. "The Sustainability Footprint of Institutional Investors." Swiss Finance Institute Research Paper 17-05.

Green, J., J. R. Hand, and X. F. Zhang. 2017. "The Characteristics That Provide Independent Information about Average US Monthly Stock Returns." *Review of Financial Studies* 30 (12): 4389–436.

GSIA. 2016. "Global Sustainable Investment Review 2016." Global Sustainable Investment Alliance. https://bit.ly/2U1OuoF.

Hasbrouck, J. 2009. "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data." *Journal of Finance* 64 (3): 1445–77.

Hoepner, A. G., I. Oikonomou, Z. Sautner, L. T. Starks, and X. Zhou. 2019. "ESG Shareholder Engagement and Downside Risk." Working paper.

Hong, H., and L. Kostovetsky. 2012. "Red and Blue Investing: Values and Finance." *Journal of Financial Economics* 103 (1): 1–19.

Jacobs, H., and S. Müller. 2020. "Anomalies across the Globe: Once Public, No Longer Existent?" *Journal of Financial Economics* 135 (1): 213–30.

Jegadeesh, N., and S. Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance* 48 (1): 65–91.

Kogan, L., and T. Wang. 2003. "A Simple Theory of Asset Pricing under Model Uncertainty." MIT working paper.

Krüger, P. 2015. "Corporate Goodness and Shareholder Wealth." *Journal of Financial Economics* 115 (2): 304–29.

Liang, H., and L. Renneboog. 2017. "On the Foundations of Corporate Social Responsibility." *Journal of Finance* 72 (2): 853–910.

Lins, K. V., H. Servaes, and A. Tamayo. 2017. "Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis." *Journal of Finance* 72 (4): 1785–824.

Lopez-de-Silanes, F., J. A. McCahery, and P. C. Pudschedl. 2019. "ESG Performance and Disclosure: A Cross-Country Analysis." ECGI working paper.

Matos, P. 2020. *ESG and Responsible Institutional Investing around the World: A Critical Review.* Charlottesville, VA: CFA Institute Research Foundation (May).

Mackintosh, J. 2018. "Is Tesla or Exxon More Sustainable? It Depends Whom You Ask." *Wall Street Journal* (17 September). https://on.wsj.com/2MQCC4m.

McLean, R. D., and J. Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *Journal of Finance* 71 (1): 5–32.

Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55 (3): 703–08.

Novy-Marx, R. 2013. "The Other Side of Value: The Gross Profitability Premium." *Journal of Financial Economics* 108 (1): 1–28.

Novy-Marx, R., and M. Velikov. 2016. "A Taxonomy of Anomalies and Their Trading Costs." *Review of Financial Studies* 29 (1): 104–47.

PRI. 2018. "Annual Report 2018." Principles for Responsible Investment. https://bit.ly/2NsU8xv.

Roll, R. 1984. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance* 39 (4): 1127–39.

USSIF. 2020. "Report on U.S. Sustainable and Impact Investing Trends 2020." Washington, DC: Forum for Sustainable and Responsible Investment.

Viale, A. M., L. Garcia-Feijoo, and A. Giannetti. 2014. "Safety First, Learning under Ambiguity, and the Cross-Section of Returns." *Review of Asset Pricing Studies* 4 (1): 118–59.

Wigglesworth, R. 2018. "Rating Agencies Using Green Criteria Suffer from 'Inherent Biases.'" *Financial Times* (20 July). https://www.ft.com/content/a5e02050-8ac6-11e8-bf9e-8771d5404543.

Wong, C., A. Brackley, and E. Petroy. 2019. "Rate the Raters 2019: Expert Views on ESG Ratings." Sustainability. https://www.sustainability.com/thinking/rate-raters-2019/.