Quantifying the Returns of ESG Investing: An Empirical Analysis with Six ESG Metrics

Florian Berg, Andrew W. Lo, Roberto Rigobon, Manish Singh, and Ruixun Zhang

Florian Berg
is a research scientist at MIT Sloan School of Management in Cambridge, MA.
fberg@mit.edu

Andrew W. Lo
is the Charles E. and Susan T. Harris Professor at the MIT Sloan School of Management; director of the MIT Laboratory for Financial Engineering; and a principal investigator at the MIT Computer Science and Artificial Intelligence Laboratory in Cambridge, MA; and an external faculty member at Santa Fe Institute in Santa Fe, NM.
alo-admin@mit.edu

Roberto Rigobon
is the Society of Sloan Fellows Professor of Management and a professor of applied economics at the MIT Sloan School of Management in Cambridge, MA.
rigobon@mit.edu

Manish Singh
is a PhD student in the MIT Department of Electrical Engineering and Computer Science in Cambridge, MA.
smanish@mit.edu

Ruixun Zhang
is an assistant professor at Peking University School of Mathematical Sciences, Center for Statistical Science, and National Engineering Laboratory for Big Data Analysis and Applications in Beijing, PRC.
zhangruixun@pku.edu.cn

KEY FINDINGS

- The authors propose methods of aggregating ESG scores from different rating agencies, addressing measurement errors and yielding aggregate measures of ESG.
- Empirically, they find significant ESG excess returns in the United States and Japan and that portfolios created using aggregate scores yield higher excess returns than portfolios constructed using individual scores.
- The authors evaluate the properties of ESG portfolios by investigating their exposure to various risk factors, constructing optimal Treynor–Black-weighted portfolios, and combining them optimally with passive index portfolios.

ABSTRACT

Within the contemporary context of environmental, social, and governance (ESG) investing principles, the authors explore the risk–reward characteristics of portfolios in the United States, Europe, and Japan constructed using the foundational tenets of Markowitz's modern portfolio theory with data from six major ESG rating agencies. They document statistically significant excess returns in ESG portfolios from 2014 to 2020 in the United States and Japan. They propose several statistical and voting-based methods to aggregate individual ESG ratings, the latter based on the theory of social choice. They find that aggregating individual ESG ratings improves portfolio performance. In addition, the authors find that a portfolio based on Treynor–Black weights further improves the performance of ESG portfolios. Overall, these results suggest that significant signals in ESG rating scores can enhance portfolio construction despite their noisy nature.

In this article, we explore the implications of environmental, social, and governance (ESG) investing principles for portfolios constructed according to the foundational tenets of modern portfolio theory as pioneered by Harry Markowitz. As we honor Markowitz's enormous impact on the field of finance, it is fitting to explore how his concepts of diversification and portfolio optimization—developed over half a century ago—still resonate within the contemporary context of ESG investing, a domain increasingly recognized for its significance in shaping sustainable financial practices. The market for ESG investing is currently estimated at $9 trillion in the United States,[1] while the number of organizations that are signatories of the United Nations Principles

[1] Source: https://www.rockpa.org/guide/impact-investing-introduction/.

for Responsible Investment (UN-PRI) has increased from 450 in 2016 to 4,935 in 2022, representing over $100 trillion of assets under management. Through a rigorous empirical analysis of ESG portfolio performances across the United States, Europe, and Japan, using data sourced from six leading ESG rating agencies over the period from 2014 to 2020, we seek to illuminate the financial efficacy of integrating ESG considerations into more traditional investment strategies. By marrying traditional financial theory with the emergent paradigms of socially responsible investing, we aim to contribute to the ongoing dialogue within this special issue, celebrating Markowitz's legacy while navigating the complex landscape of the financial industry toward more ethical and sustainable horizons.

Investors who want to implement ESG factors in their portfolios typically rely on ESG scores provided by third-party rating agencies that specialize in measuring ESG performance. There is a rapidly growing number of such rating agencies, and in our sample, we rely on ratings from some of the largest, including MSCI Inc, S&P Global, ISS, Moody's ESG Solutions, Reprisk, and Truvalue Labs, each of which has a proprietary methodology for the calculation of their ratings. The scoring process typically involves gathering data from a variety of sources, including yearly regulatory filings, media reports, and self-disclosed data from firms and international organizations. The choice of different data methodologies and sources can lead to a substantial divergence between rating providers (Berg, Kölbel, and Rigobon 2022).[2] ESG ratings, and sustainable investing in general, have received a number of high-profile critiques. For example, *The Economist* dedicated a recent cover story to ESG investing,[3] concluding that ESG ratings are too complex and contain too much measurement error to be useful. This raises the question of whether ESG ratings are, in fact, useful for portfolio construction and, if so, how to optimally exploit the signals in ESG ratings, despite their noisy nature.

One of the most prominent critiques of ESG ratings raised by regulators concerns the fiduciary responsibility of financial institutions. Portfolios constructed using ESG scores are constrained, and, therefore, if such a constraint reduces portfolio returns, it could be considered a breach of fiduciary duty—especially if the financial institution fails to clearly inform the investor of such a possibility. In general, most of the regulatory commentary is based on the intuition that restricted portfolios are, by necessity, less profitable than unrestricted ones. However, this intuition is correct only if a constraint is orthogonal to returns. That is not the case if the selection mechanism is associated with the fundamental characteristics of the stocks. Consequently, understanding whether ESG scores are associated with excess returns is of crucial importance to investors, regulators, financial institutions, and ultimately, to those that care about ESG impact on society in general.

In our empirical analysis, we construct ESG portfolios for the US, European, and Japanese stock markets, using ESG scores from six major rating agencies from 2014 to 2020. We quantify the excess returns of these portfolios with respect to standard asset pricing models, including the capital asset pricing model (CAPM) and Fama–French factor models. We find a wide range of excess returns in portfolios constructed using different ESG scores. For example, the MSCI-based portfolio that goes long the top quartile of stocks based on ESG rating and short the bottom quartile achieves a statistically significant annual alpha of 3.8% in excess of the Fama–French five-factor model in the United States, while the same portfolio using ESG ratings from other agencies shows much lower (and usually neutral) excess returns. In addition,

[2] Furthermore, Berg et al. (2021) show that these ratings contain a considerable amount of measurement error.

[3] "ESG Investing: A Broken Idea." The Economist, Special Report (July 23, 2022): https://www.economist.com/special-report/2022-07-23.

the same rating agency may have very different excess returns across regions. This instability in coefficients is to be expected, considering the sizeable noise in ESG scores.

To address the problem of noise, we propose several different ways to aggregate ESG scores across vendors. We construct a single measure by combining individual ESG scores from six different vendors using various statistical and voting aggregation techniques, including simple averages, the Mahalanobis distance, principal component analysis (PCA), average voting, and singular transferable voting. Our goal is to retain the ESG signal in the aggregate rating while attenuating the noise. Different aggregation methods will necessarily weight the ESG scores from rating agencies differently. For example, the simple average attributes equal weights to scores from all vendors, while the Mahalanobis distance aggregates ratings based on their variance–covariance, and PCA weights the rating agencies in such a way to retain the direction of their maximum observed variance.

We find that aggregating individual ESG ratings improves portfolio performance significantly. We construct sorted ESG portfolios (from high to low scores) and analyze their risk-adjusted returns, excess returns, and exposures to fundamental factors. In particular, we find that portfolios in the United States based on the Mahalanobis distance achieve the highest annualized alpha, over 6%, while portfolios based on singular transferable voting achieve the highest annualized alpha in Europe (over 6%) and Japan (over 9%).

The empirical evidence on excess returns of ESG investing has been mixed in the existing literature. Some document a positive relationship between ESG scores and excess returns (see, for example, Edmans 2011; Khan, Serafeim, and Yoon 2016; Lins, Servaes, and Tamayo 2017; Albuquerque, Koskinen, and Zhang 2019), while others find a negative relationship (see, for example, Chava 2014; El Ghoul et al. 2011; Bolton and Kacperczyk 2020). Berg et al. (2021) describe a theoretical model explaining both relationships. They find that positive realized returns are explained by unexpected inflows into stocks with high ESG performance, and as these inflows level out, the expected returns become lower. Put differently, high ESG-rated firms benefit from a lower cost of capital due to higher market capitalization. Another possible explanation is found in omitted variable bias, in particular the omission of management quality. If good ESG performance is correlated to high management quality, the link between ESG performance and returns would no longer be causal. In our sample from 2014 to 2020, we found a positive relationship in the United States and Japan, most likely due to inflows of funds from new ESG investors into high ESG-rated stocks. For example, Berg, Heeb, and Kölbel (2022) find that MSCI rating changes drive changes in ESG mutual fund holdings in the US market, albeit with a very slow integration of up to 18 months. They also show that this correlates temporally with returns.

We find that the portfolio construction methodology proposed by Lo and Zhang (2023) further improves the performance of ESG portfolios. Lo and Zhang's (2023) methodology begins by quantifying the excess returns for individual assets using a small number of parameters,[4] and then it applies Treynor–Black weights to optimize the Sharpe ratio of an ESG portfolio in which the weights are proportional to the rank of the ESG score of each firm.[5] Using this framework, we achieve improved excess returns in ESG portfolios, especially for portfolios with a large number of assets.

[4] Specifically, it uses the cross-sectional correlation between ESG scores and excess returns of each stock.

[5] We compare the excess returns of ESG portfolios using the model of Lo and Zhang (2023) with their forward-looking realized excess returns and find a high degree of consistency between the two, thereby validating the model.

This is valid in particular for portfolios constructed to go long the stocks in the top four deciles and short the stocks in the bottom four deciles, so that weights based on the rank of each firm's ESG score have a meaningful impact.

Some investors may prefer to rely on E, S, or G scores individually in the creation of their portfolios. Consequently, we also investigate the aggregation of individual E, S, and G scores across vendors, and analyze the excess returns of top–bottom sorted portfolios. We find the highest excess returns for portfolios based on E scores in the United States and Japan. In portfolios based on S and G scores, we find positive excess returns only for some portfolios and aggregation methods.

Our article is related to several strands in the literature. The first strand is about disagreement between ESG providers. Our work is based on the growing literature that highlights the divergence between ESG ratings (see, for example, Dorfleitner, Halbritter, and Nguyen 2015; Semenova and Hassel 2015; Berg, Kölbel, and Rigobon 2022; Brandon, Krueger, and Schmidt 2021).

The second strand explores the relationship between ESG and stock returns. Some research shows higher returns (Edmans 2011; Khan, Serafeim, and Yoon 2016; Lins, Servaes, and Tamayo 2017), while other work shows a negative relationship both empirically (Bolton and Kacperczyk 2020) and theoretically (Pástor, Stambaugh, and Taylor 2021). Pástor, Stambaugh, and Taylor (2022) show that the high returns for green assets in recent years reflect unexpectedly strong increases in environmental concerns, not high expected returns. Our work differs in our acknowledgement of the noisiness of ESG ratings and our proposal of different aggregation methods.

Finally, the third strand is related to the nascent literature in dealing with measurement noise in ESG ratings and its impact on returns. Berg et al. (2021) use instrumented variable regressions to remove the noise in one version of the ESG score using others. In our work, we improve the signal using aggregation methods that combine multiple sources of data, and we leverage the optimal portfolios of Lo and Zhang (2023) to further improve the performance of these ESG portfolios.

METHODOLOGY

In this section, we describe the methodology used to construct portfolios based on ESG scores. We discuss our strategy to quantify excess returns for individual stocks, the portfolio construction methodologies, and several methods to aggregate multiple ESG scores.

Quantifying Excess Returns

We start by describing a methodology first proposed by Lo and Zhang (2023), which we adapt to ESG portfolios. We quantify the excess returns (alphas) of individual stocks ranked by their ESG scores. This allows us to optimize the weights used in ESG portfolios and to quantify their portfolio returns.

We consider a universe of N stocks with returns, R_{it}, that satisfy the following linear multifactor model (e.g., the Fama–French factor model):

$$R_{it} - R_{ft} = \alpha_i + \beta_{i1}(\Lambda_{1t} - R_{ft}) + \cdots + \beta_{iK}(\Lambda_{Kt} - R_{ft}) + \epsilon_{it} \tag{1}$$

$$\text{such that } \mathbb{E}[\epsilon_{it} | \Lambda_{kt}] = 0, \quad k = 1, \ldots, K, \tag{2}$$

where Λ_{kt} is the k-th factor return, $k = 1, \ldots, K$; R_{ft} is the risk-free rate; α_i and β_{ik} are the excess returns and factor betas, respectively; and ϵ_{it} is the idiosyncratic return component.

ESG investors typically rank stocks according to their ESG scores, which we denote by ESG_i, and we use $\alpha_{[i:N]}$ to represent the alpha of the i-th ranked stock.[6] Lo and Zhang (2023) show that the expected values, variances, and covariances of these ranked alphas are given by

$$\mathbb{E}(\alpha_{[i:N]}) = \sigma_\alpha \cdot \rho \cdot \mathbb{E}(Y_{i:N}), \tag{3}$$

$$\mathrm{Var}(\alpha_{[i:N]}) = \sigma_\alpha^2 \cdot (1 - \rho^2 + \rho^2 \cdot \mathrm{Var}(Y_{i:N})), \tag{4}$$

$$\mathrm{Cov}(\alpha_{[i:N]}, \alpha_{[j:N]}) = \sigma_\alpha^2 \cdot \rho^2 \cdot \mathrm{Cov}(Y_{i:N}, Y_{j:N}), \tag{5}$$

for i, $j = 1, 2, \ldots, N$, and $i \neq j$. Here, ρ is the cross-sectional correlation between α_i and ESG_i;[7] σ_α is the standard deviation of α_i; and $Y_{1:N} < Y_{2:N} < \cdots < Y_{N:N}$ are the order statistics of N independent and identically distributed standard Gaussian random variables.

Portfolio Creation

Given a set of ESG scores for all firms, we construct ESG portfolios and estimate their performances. At time t, we sort the stocks based on ESG scores and construct three long–short portfolios. The first, $pf_{(\pm 10)}$, represents a portfolio that goes long the top decile of stocks with equal weights (denoted by $pf_{(+10)}$) and shorts the bottom decile of stocks with equal weights (denoted by $pf_{(-10)}$). The second, $pf_{(\pm 25)}$, represents a portfolio that goes long the top quartile of stocks with equal weights (denoted by $pf_{(+25)}$) and shorts the bottom quartile of stocks with equal weights (denoted by $pf_{(-25)}$). The third, $pf_{(\pm 40)}$, represents a portfolio that goes long the top four deciles of stocks with equal weights (denoted by $pf_{(+40)}$) and shorts the bottom four deciles of stocks with equal weights (denoted by $pf_{(-40)}$). The total weights of individual stocks on both the long and short sides are set to one to give all portfolios the same amount of leverage. We henceforth refer to these three portfolios as the ±40, ±25, and ±10 portfolios. These portfolios are rebalanced once a year because ESG scores are updated by some vendors once a year. We observed very high cross-sectional autocorrelations between scores, as shown in Exhibit A1 in the Appendix.

In addition to these equal-weighted portfolios, we build optimized Treynor–Black portfolios using the model-implied alphas for individual stocks in Equations (3)–(5), which use the rank of stocks in the ESG-sorted portfolio. For example, in $pf_{(+10)}$ portfolios, all stocks in the 10th decile (percentiles 90 to 100) are given equal weights. However, in Treynor–Black weighting, higher-ranked stocks are given larger weights than lower-ranked stocks, if the correlation, ρ, between the ESG score and stock alpha is positive. Specifically, Lo and Zhang (2023) show that the weight of

[6] In the statistics literature, these indirectly ranked variables are termed *induced order statistics* (Bhattacharya 1974), because they are ranked not by their own values (α_i in our case) but by the values of another variable (ESG_i in our case). As such, α_i are modeled as random variables to reflect the fact that they may be correlated with the ESG scores. This specification was used in Lo and MacKinlay (1990) to represent the cross-sectional estimation errors of intercepts derived from CAPM regressions. In the current context, we interpret the randomness in α_i as a measure of uncertainty regarding the degree of mispricings of stocks, which is similar to the treatment in Pástor and Stambaugh (1999).

[7] Lo and Zhang (2023) assume that they are jointly normally distributed, and Lo et al. (2024) generalize the framework to arbitrary marginal distributions.

the i-th ranked stock in a universe of N stocks can be approximated by the following equation:[8]

$$\omega_i \propto \Phi^{-1}(\zeta_i), \tag{6}$$

where $\zeta_i = i/N$ and Φ^{-1} is the inverse of the cumulative standard normal distribution.

Once these ESG portfolios are constructed, we can further combine them with any other portfolio. The most natural application is to combine the active (ESG) portfolio with a passive index such as the market portfolio. The returns of the combined portfolio are given by

$$r_{active+passive} = \omega_A * r_{active} + (1 - \omega_A) * r_{passive}, \tag{7}$$

where r_{active} can be any return of a top–bottom equal-weighted or Treynor–Black weighted portfolio, $r_{passive}$ is the return of the passive portfolio (e.g., the market index), and ω is the weight of the active portfolio. In our analysis, we have fixed ω_A at 0.5 as an illustrative example.

In a later section, Performance of ESG Portfolios, we will evaluate the performance of all equal-weighted, Treynor–Black, and active-plus-passive portfolios.

ESG Rating Aggregation

Berg et al. (2021) show that ESG ratings are certainly noisy but nevertheless contain a signal. The measurement error inherent to ESG ratings makes it difficult to find significant risk premia. To solve this problem, we propose several different ways of aggregating ESG scores. Let $ESG_{i,t,j}$ be the ESG score of company i rated by vendor j at time t. We then compute $ESG_{i,t,m}$, where m is the aggregation method. We describe each aggregation method next and analyze the performance of portfolios based on these aggregate ESG scores in a later section.

Equal-Weighted Average

The first and simplest aggregation method under consideration is the equal-weighted cross-sectional average of all ESG ratings. The average is a widely used method for noise attenuation, and intuitively, we expect the ESG signal to be stronger here than in the case of a single rating. We use *AVG* to indicate the equal-weighted average rating.

$$ESG_{i,t,avg} = \sum_{j \in \{MSCI, S\&PGlobal, \ldots, TVL\}} ESG_{i,t,j} \tag{8}$$

Principal Component Analysis

PCA is a widely used dimensionality reduction method. In some cases, it can also be used as a tool for noise reduction. The PCA performs a change of basis transformation and projects the data in the direction of the maximum variance. If errors are similarly distributed between ratings, it will minimize the information loss. We treat the ESG scores from the six vendors in our sample as high-dimensional data and obtain its lower dimensional (1-d) representation as the aggregate score.

$$ESG_{i,t,PCA} = PCA(ESG_{i,t,MSCI}, ESG_{i,t,S\&PGlobal} \ldots ESG_{i,t,TVL})$$

[8] The approximation holds when the number of stocks is large and stocks have identical idiosyncratic volatilities.

We use *PCA* to indicate the aggregate score obtained by principal component analysis.

Mahalanobis Distance

The Mahalanobis distance is the distance between two points after accounting for variances and covariances across dimensions. In brief, highly correlated ESG scores will not be given excess weight in the calculation of the aggregate score. Let the ESG scores of firm *i* be the vector *x* and *y* be the vector with minimum ESG scores. We then calculate the Mahalanobis distance as follows:

$$\begin{aligned} x_{i,t} &= (ESG_{i,t,MSCI}, ESG_{i,t,S\&PGlobal} \ldots ESG_{i,t,TVL}) \\ y_t &= (min_i(ESG_{i,t,MSCI}), min_i(ESG_{i,t,S\&PGlobal}) \ldots min_i(ESG_{i,t,TVL})) \\ ESG_{i,t,Maha} &= \sqrt{(x_{i,t} - y_t)^T S_t^{-1} (x_{i,t} - y_t)}, \end{aligned} \tag{9}$$

where S_t is the variance–covariance matrix for the ESG ratings at time *t*. We refer to the aggregate score obtained by the computation of the Mahalanobis distance as *MAHA*.

Voting Average

The voting average is based on the theory of social choice, that is, the aggregation or combination of individual preferences in collective decisions. In voting aggregation methods, the ESG scores from a rating agency are considered as a ranked list of choices and the different agencies are considered as voters. In the voting average process, each rank is averaged to compute an aggregate rank of each firm. It is represented by AVG_{vote} in our subsequent analysis.

Singular Transferable Voting

Using the singular transferable vote (STV) method to aggregate ESG ratings, the least preferred candidate is eliminated and the vote is transferred to the next preferred candidate. The process is repeated until all the candidates are eliminated, one by one, forming a ranked list of candidates based on the order of elimination. The aggregate ESG score using this method is represented by STV_{vote} in our subsequent analysis.

Optimized ESG Score

The optimized ESG score is a linear combination of ESG scores from all rating agencies. The weights are derived from an optimization that maximizes the cross-sectional correlation between ESG scores and the one-year future excess returns of the stocks.

$$\begin{aligned} ESG_{i,t,opt} &= \sum_{r \in MSCI, S\&PGlobal \ldots} w_r * ESG_{i,t,r} \\ & w_r : max(avg_t(corr_i(ESG_{i,t,opt}, \alpha_{i,t}))) \end{aligned} \tag{10}$$

Because the optimization involves the use of out-of-sample excess returns, the excess returns obtained using $ESG_{i,t,opt}$ cannot be realized. However, the ESG scores optimized in this way have the maximum correlation with excess returns that can be achieved by a linear aggregation of ESG scores from different rating agencies. The aggregate scores obtained using optimization are referred to as *OPT* in our analysis.

DATA AND SUMMARY STATISTICS

We obtained data from six leading ESG rating providers, including MSCI, S&P Global, ISS, Moody's ESG Solutions, Reprisk, and Truvalue Labs.[9] We observe non-overlapping coverage in our dataset across rating agencies, so for a fair comparison across different providers, we include only those firms with observations from all six rating agencies. We also have E, S, and G ratings in our dataset for all agencies, except for Truvalue Labs, which does not offer such scores.

For our analysis, we classify firms into three different regions: the United States, Europe, and Japan. The total number of firms in each region are 633 in the United States, 547 in Europe, and 274 in Japan. Our sample spans from March 2014 to 2020. The relatively short time series is explained by the fact that sustainable investing is a fairly recent phenomenon. Because ESG ratings from different providers have different scales, we renormalize them to have zero mean and unit variance in the cross section.

The daily returns were queried from the Refinitiv Workspace.[10] For the calculation of excess returns, we obtained our data from the Fama–French data library.[11] We retrieved the daily MSCI index returns for the United States, Europe, and Japan from MSCI Index Solutions.[12]

Summary Statistics

We averaged the cross-sectional rank correlations between different individual and aggregate ESG scores over time and present the results in Exhibit 1.[13] The correlation between different ESG ratings varies over time, and some scores are highly correlated with each other, such as ISS, MSCI, S&P Global, and Moody's. Rather surprisingly, Reprisk has a negative correlation with the other ratings.

EXHIBIT 1

The Cross-Sectional Rank Correlation between ESG Scores (individual and aggregate) Averaged over Time

	ISS: ESG	MSCI: ESG	Reprisk: ESG	S&P Global: ESG	TVL: ESG	Moody's: ESG	AVG: ESG	PCA: ESG	MAHA: ESG	AVG_{vote}: ESG	STV_{vote}: ESG
ISS: ESG	1.00	0.40	–0.24	0.59	0.12	0.62	0.73	0.80	0.48	0.75	0.59
MSCI: ESG	0.40	1.00	–0.02	0.34	0.18	0.38	0.67	0.57	0.45	0.68	0.33
Reprisk: ESG	–0.24	–0.02	1.00	–0.41	0.14	–0.33	0.03	–0.43	0.43	0.00	–0.14
S&P Global: ESG	0.59	0.34	–0.41	1.00	0.06	0.69	0.67	0.84	0.46	0.69	0.60
TVL: ESG	0.12	0.18	0.14	0.06	1.00	0.10	0.46	0.18	0.45	0.45	0.56
Moody's: ESG	0.62	0.38	–0.33	0.69	0.10	1.00	0.73	0.85	0.46	0.74	0.50
AVG: ESG	0.73	0.67	0.03	0.67	0.46	0.73	1.00	0.84	0.84	0.99	0.73
PCA: ESG	0.80	0.57	–0.43	0.84	0.18	0.85	0.84	1.00	0.50	0.86	0.66
MAHA: ESG	0.48	0.45	0.43	0.46	0.45	0.46	0.84	0.50	1.00	0.80	0.62
AVG_{vote}: ESG	0.75	0.68	0.00	0.69	0.45	0.74	0.99	0.86	0.80	1.00	0.72
STV_{vote}: ESG	0.59	0.33	–0.14	0.60	0.56	0.50	0.73	0.66	0.62	0.72	1.00

[9] We also have data from Sustainalytics and Refinitiv, but both have changed their methodologies over time, and their ESG scores were backfilled using these new methodologies (Berg, Fabisik, and Sautner 2021). Adding these two providers to our analysis would introduce a forward-looking bias.

[10] Refinitiv Workspace has since been rebranded and it is now known as LSEG Workspace. See https://www.lseg.com/en/data-analytics/products/workspace.

[11] Available online at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[12] See https://www.msci.com/our-solutions/indexes.

[13] In the construction of a long–short ESG portfolio, the rank of a company matters more than its score value.

EXHIBIT 2

Alpha Statistics

	FF5		FF3		CAPM	
Region	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
United States	1.02	19.76	1.13	19.98	–0.55	22.93
Europe	6.55	21.46	6.14	21.60	3.86	23.35
Japan	4.27	18.64	3.29	19.13	0.14	20.72

NOTE: The cross-sectional mean and standard deviation (Std) of excess returns (alpha) of the sampled companies, computed using different factor models—Fama–French five-factor (FF5), Fama–French three-factor (FF3), and CAPM—averaged over time.

Among the aggregate scores, AVG, PCA, and AVG_{vote} exhibit relatively high correlations (>80%). However, the aggregate scores based on the STV method are significantly different from other aggregate scores, given their smaller correlation coefficients (60%–70%). We find relatively similar correlations (in the range of 40%–50%) between individual ESG scores and scores computed using the Mahalanobis distance. This is due to the variance–covariance normalization property of the Mahalanobis distance.

Excess Returns

In Exhibit 2, we present the cross-sectional mean and standard deviation of excess returns (alphas, $\alpha_{i,t}$: the one-year future alpha of the stock i at time t) of the sampled companies by region, computed as follows: $mean_\alpha = avg_t(avg_i(\alpha_{i,t}))$, $std_\alpha = avg_t(std_i(\alpha_{i,t}))$, where i is the company and t is the time. We find that the cross-sectional standard deviation of excess returns varies between 18% and 25%, while the cross-sectional mean of excess returns varies widely between regions. These parameters are necessary to quantify the excess returns of portfolios constructed as described in the earlier Methodology section.

PERFORMANCE OF ESG PORTFOLIOS

In this section, we begin our study of the performance of ESG portfolios by first measuring the cross-sectional correlation between ESG scores and excess returns of individual stocks. We next summarize the empirical properties of both the raw returns and the returns in excess of Fama–French factor models for portfolios based on their individual and aggregate ESG scores. We then look at ESG portfolios constructed using Treynor–Black weights and ESG portfolios combined with passive index portfolios.

Correlations between ESG Ratings and Excess Returns

Given the ESG scores of individual stocks and their excess returns, we compute the cross-sectional correlations, $corr_t = corr_i(ESG_{i,t}, \alpha_{i,t})$, on date t for different ESG scoring methods, where $\alpha_{i,t}$ is the one-year forward-looking alpha of stock i at date t. The average values of $corr_t$ are presented in Exhibit 3 for both individual and aggregate ESG scores. The last row shows that the maximum possible correlation (OPT: ESG) from linear combinations of individual ESG scores is in the range of 6% to 10% for all alphas and regions. Although the correlations for other aggregate ESG scores are inevitably lower, they are not much lower compared with the optimized ESG score. Different regions have different scores with the highest correlations. For example, in the United States, MAHA: ESG and AVG: ESG achieve the highest correlations, while ISS: ESG has the highest correlations in Europe.

In Exhibit 4, we show the time series correlation values for AVG: ESG scores. We observe that correlations are typically in the range of 30% to 40% in the United States and Japan. However, there are some instances of negative correlation as well. A negative correlation implies that a top–bottom portfolio has negative excess returns. We find positive correlations in our sample, on average, because from 2014 to 2020, the ESG portfolio potentially had positive excess returns.

EXHIBIT 3

Average Cross-Sectional Correlation between ESG Scores and Alpha (excess returns)

	United States			Europe			Japan		
ESG	FF5	FF3	CAPM	FF5	FF3	CAPM	FF5	FF3	CAPM
ISS: ESG	2.88	3.40	7.75	5.66	7.03	4.23	6.07	6.86	8.43
MSCI: ESG	4.46	4.77	6.70	3.74	5.13	5.38	5.41	5.09	6.23
Reprisk: ESG	2.75	2.36	2.71	–3.22	–1.37	6.11	2.40	1.88	4.45
S&P Global: ESG	1.72	1.90	2.58	3.63	2.83	–3.66	6.41	5.51	6.01
TVL: ESG	2.17	2.92	3.43	2.42	4.43	5.62	3.63	3.47	2.24
Moody's: ESG	2.81	2.82	4.38	3.22	3.12	–2.60	5.66	5.33	5.02
AVG: ESG	4.91	5.32	8.09	4.70	6.41	4.61	8.74	8.32	9.57
PCA: ESG	2.79	3.20	5.55	5.33	5.51	–0.36	6.73	6.57	6.93
MAHA: ESG	5.42	5.52	7.88	1.78	3.62	4.96	9.02	8.26	9.91
AVG_{vote}: ESG	4.90	5.40	7.67	5.11	6.69	4.62	8.85	8.66	9.89
STV_{vote}: ESG	3.59	4.18	5.45	4.31	4.86	0.86	8.25	7.85	7.74
OPT: ESG	6.06	6.14	9.32	6.56	7.99	5.32	9.13	8.47	10.37

NOTE: FF3: Fama–French three-factor model; FF5: Fama–French five-factor model; TVL: Truvalue Labs.

The different statistics displayed in Exhibits 2–3 can be used to obtain the average excess returns as described in the earlier Quantifying Excess Returns subsection (in the Methodology section). For example, the average alpha for the ±10 ESG portfolio with a cross-sectional α deviation 20% and correlation value of 5% will be $2 \times 0.20 \times 5 \times 1.64 = 3.28\%$. We discuss the excess returns of different ESG portfolios in more detail later in this section.

Returns and Factor Exposure

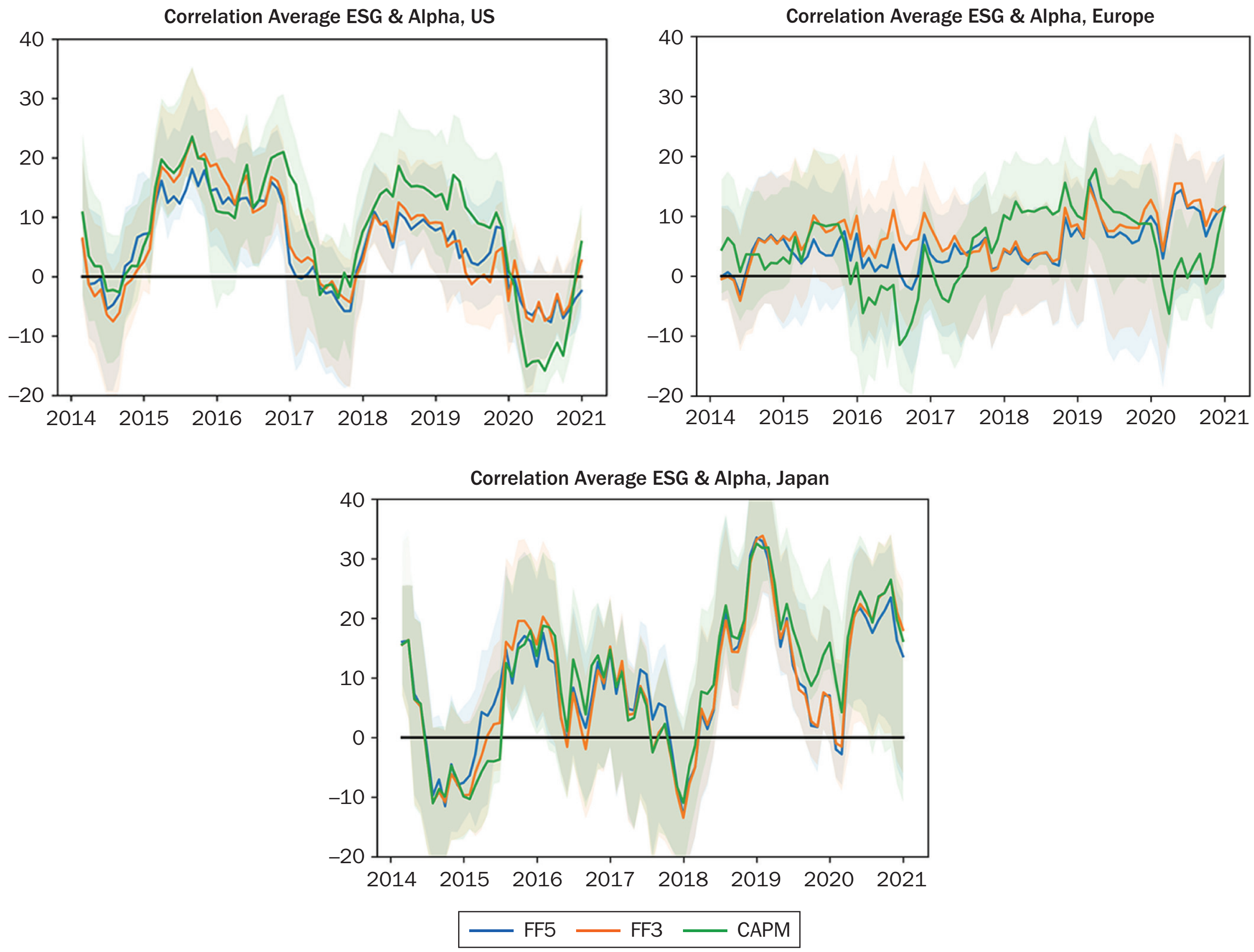
Following the results in the earlier Portfolio Creation subsection (in the Methodology section), we construct multiple ESG top–bottom portfolios (±40, ±25, and ±10) and compute their properties, such as their returns, risk-adjusted returns (Sharpe ratios), excess returns (alphas), and exposure to different factors. In addition to portfolios based on the ESG scores from individual vendors, we also construct portfolios using the aggregate scores computed with the methods described in the earlier subsection ESG Rating Aggregation in the Methodology section. We also include a portfolio that is simply the average portfolio constructed using individual ESG scores (indicated as INDI-AVG: ESG).[14]

Exhibit 5 presents the mean returns and annualized Sharpe ratios for ESG portfolios computed using individual and aggregate ESG scores. The ESG portfolios achieve annualized returns as high as 8% in the United States and Japan and 6.3% in Europe. Similarly, the highest Sharpe ratios are more than 1 in the United States and Japan (up to 1.23) and 0.96 in Europe. We find that portfolios based on ESG scores from individual vendors have nonuniform returns and Sharpe ratios. ESG scores from ISS, MSCI, and Reprisk have consistently positive returns across different regions, while portfolios based on scores from S&P Global, Truvalue Labs, and Moody's have negative returns in some regions. However, the portfolios using aggregate scores have positive returns, and the returns are generally higher than those of portfolios based on individual vendor scores.

[14] This portfolio differs from the AVG: ESG portfolio because in AVG: ESG, we first compute the average ESG scores and then construct portfolios based on the average. In INDI-AVG: ESG, however, the first step is to construct portfolios based on individual ESG scores and the second step is to average the returns of individual ESG portfolios.

EXHIBIT 4

The Correlation between Average ESG Score and Stock Performance (alpha) versus Time (year)



NOTE: We find that the correlation between α and the average ESG score can be as high as 20% to 30% (in the United States and Japan, respectively); however, it clearly varies over time.

Given the time series of portfolio returns, we follow Equation (1) and perform a time series regression to compute the excess returns and fundamental factor exposure of these ESG portfolios. The response variable of the regression is the portfolio return, and the factors are the Fama–French five factors.

We have a total of 33 ESG portfolios (11 different ESG scores multiplied by 3 top–bottom portfolios) and find that the portfolios have varying exposures to multiple fundamental factors. We present the exposures in Exhibit 6, which gives the number of positive and negative significant betas (defined by a p-value < 5%) out of 33 ESG portfolios. In the United States and Europe, ESG portfolios tend to have a negative exposure to the market and size factors and a positive exposure to the profitability factor. However, in Japan, ESG portfolios tend to have negative exposure to the profitability and investment factors.

A varying exposure to different risk factors implies that the returns of the portfolios are not purely due to ESG risk premia. They may be due to exposure to different fundamental factors, as shown by the number of statistically significant coefficients (positive and negative) in Exhibit 6. Therefore, we next analyze the excess returns of the ESG portfolios.

EXHIBIT 5

Mean Returns and Sharpe Ratios for the Top *y*% minus Bottom *y*% ESG Portfolios (individual and aggregate scores) for the United States, Europe, and Japan, where *y* = 40, 25, 10

	United States			Europe			Japan		
Portfolio	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**
Mean Returns									
ISS: ESG	4.89	4.57	5.51	2.37	1.66	1.10	4.68	5.76	6.98
MSCI: ESG	3.25	4.72	5.48	3.89	4.13	3.76	3.48	5.24	1.53
Reprisk: ESG	2.07	2.05	4.17	1.21	3.29	6.30	1.14	2.31	4.29
S&P Global: ESG	–0.08	–0.48	1.46	–1.14	–0.70	2.02	2.62	3.63	7.78
TVL: ESG	2.39	2.49	–0.46	2.00	3.05	1.14	–0.52	–0.24	1.11
Moody's: ESG	1.58	2.19	3.91	–0.93	–0.29	–1.89	2.35	1.86	6.39
INDI-AVG: ESG	2.35	2.59	3.34	1.23	1.85	2.18	2.29	3.09	4.68
AVG: ESG	3.32	6.48	6.39	2.60	3.21	4.30	5.26	5.38	4.83
PCA: ESG	2.26	3.69	2.94	0.28	0.68	1.54	3.04	3.81	4.20
MAHA: ESG	4.18	6.03	7.50	2.15	3.46	3.47	4.43	5.46	5.52
AVG_{vote}: ESG	3.18	5.56	6.85	2.29	3.49	4.46	4.99	6.13	5.29
STV_{vote}: ESG	2.59	2.07	1.30	–0.25	1.34	5.75	2.97	4.32	8.00
OPT: ESG	4.77	6.54	8.36	1.64	2.09	5.76	4.97	5.99	5.24
Sharpe Ratio									
ISS: ESG	1.04	0.83	0.81	0.57	0.33	0.15	1.07	1.05	0.84
MSCI: ESG	1.07	1.23	0.94	1.00	0.86	0.53	0.93	1.05	0.19
Reprisk: ESG	0.53	0.45	0.67	0.27	0.57	0.73	0.25	0.39	0.51
S&P Global: ESG	–0.03	–0.12	0.26	–0.28	–0.13	0.28	0.53	0.53	0.82
TVL: ESG	0.60	0.50	–0.07	0.58	0.75	0.21	–0.13	–0.05	0.17
Moody's: ESG	0.45	0.45	0.55	–0.24	–0.06	–0.28	0.42	0.26	0.68
INDI-AVG: ESG	1.03	0.92	0.92	0.63	0.75	0.58	0.82	0.87	1.06
AVG: ESG	0.81	1.18	0.85	0.70	0.71	0.62	1.12	0.93	0.58
PCA: ESG	0.59	0.73	0.42	0.07	0.13	0.21	0.59	0.57	0.49
MAHA: ESG	0.99	1.09	1.04	0.61	0.79	0.51	1.17	1.08	0.73
AVG_{vote}: ESG	0.83	1.03	0.97	0.61	0.75	0.66	0.99	1.00	0.61
STV_{vote}: ESG	0.62	0.41	0.20	–0.07	0.31	0.96	0.67	0.70	1.03
OPT: ESG	1.14	1.25	1.22	0.43	0.43	0.79	1.15	1.07	0.67

EXHIBIT 6

Number of Significant Positive (+) and Negative (–) Betas (defined by a *p*-value <5%) for All 33 Top–Bottom ESG Portfolios (11 different ESG scores multiplied by 3 top–bottom portfolios)

		Market		Size		BM		Profitability		Investment	
		+	–	+	–	+	–	+	–	+	–
11 × 3 Portfolios	United States	0	8	3	22	4	13	4	0	1	2
	Europe	0	22	3	25	14	3	25	0	0	3
	Japan	5	0	0	2	4	9	0	17	0	18

NOTE: BM: Book-to-market ratio.

Estimating Excess Returns

In this subsection, we evaluate the returns of ESG portfolios in excess of their Fama–French factors by estimating a forward-looking time series regression of raw returns on Fama–French factors. We compute the alphas of yearly top–bottom quantile portfolios.

We report the annualized excess returns of the top–bottom ESG portfolios (±40, ±25, and ±10), as described earlier. In Exhibit 7, we present the excess returns (alphas) from the time series regression, shown in Equation (1), using the Fama–French five-factor model and the CAPM. From Exhibit 7, we see that the excess returns are positive and significant in the United States and Japan, but not in Europe. The CAPM alphas are generally higher than the Fama–French five-factor model alphas, implying that the Fama–French factors partially explain the positive returns beyond the market factor of the CAPM.

The excess returns of the OPT portfolios are comparable to the highest returns from those constructed using other aggregation methods or by individual scores. The excess returns from the ESG portfolios using OPT cannot be realized, however, but the other aggregate and individual ESG-score alphas can. Hence, the highest realizable alpha is close to the maximum possible alpha that can be obtained using a linear combination of different ESG scores.

EXHIBIT 7

Fama–French Five-Factor Model (FF5) and CAPM Alphas for Top–Bottom ESG Portfolios (±40%, ±25%, ±10%)

	United States			Europe			Japan		
Portfolio	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**
FF5 Alpha									
ISS: ESG	3.30***	2.34*	3.01*	1.95	1.82	0.27	4.27**	5.96***	7.71***
MSCI: ESG	2.47**	3.80***	4.81**	2.90**	2.89**	1.49**	3.93**	5.84***	3.59***
Reprisk: ESG	0.95	0.86	2.75	–1.46	–0.40	0.80	0.09	1.27	2.88**
S&P Global: ESG	0.72	0.31	1.92	0.43	1.22	3.26*	3.99***	5.11***	9.09***
TVL: ESG	2.37	1.79	–1.04	0.11	0.83	–1.28	–0.41	0.12	0.77
Moody's: ESG	1.59*	2.73*	4.40**	0.58	1.29	–0.48	2.71*	1.94*	7.04**
INDI-AVG: ESG	1.90**	1.98**	2.64**	0.76	1.28	0.78	2.43**	3.37**	5.19***
AVG: ESG	2.69**	5.40***	4.50**	1.82*	2.32*	2.12*	5.58***	5.52**	5.92**
PCA: ESG	1.91*	3.23**	3.36*	1.70	1.93	2.51	3.92***	4.81**	5.17**
MAHA: ESG	3.44***	4.34**	6.09***	0.34	0.95	–0.49	4.14**	5.79***	7.25*
AVG_{vote}: ESG	2.58**	4.36***	5.44**	1.60	2.82*	2.85*	5.31***	6.33***	7.00**
STV_{vote}: ESG	2.96**	2.44**	0.88**	–0.49	1.39	6.06***	3.69**	5.36**	9.56***
OPT: ESG	3.87***	4.70***	5.73**	1.24	1.52	5.20**	5.32***	6.12***	6.47**
CAPM Alpha									
ISS: ESG	5.94***	5.39***	6.27***	2.46**	1.72**	1.31**	4.18**	5.22***	6.49**
MSCI: ESG	3.51***	5.00***	6.58***	4.11***	4.42**	4.21*	3.14**	4.79**	1.22**
Reprisk: ESG	1.86**	1.48**	2.89**	1.28	3.44*	6.57**	1.27	2.32	4.35***
S&P Global: ESG	0.16	–0.26	2.26	–1.15	–0.75	1.96	2.09	2.87	6.56*
TVL: ESG	3.30**	3.02**	–0.30**	1.97	3.07*	1.14*	–1.20	–1.15	0.66
Moody's: ESG	2.14*	3.11**	5.10***	–0.92	–0.29	–1.96	1.32	0.41	4.82
INDI-AVG: ESG	2.82***	2.96**	3.80***	1.30**	1.94**	2.32*	1.79	2.40*	4.01**
AVG: ESG	4.23**	7.72***	7.41***	2.72**	3.34**	4.60**	4.58***	4.46**	3.71**
PCA: ESG	2.96**	4.76***	5.12**	0.32	0.75	1.73	2.35	2.88	3.17
MAHA: ESG	5.16***	6.89***	8.60***	2.29*	3.66**	3.80**	3.98**	4.76**	4.75**
AVG_{vote}: ESG	3.91***	6.55***	7.25***	2.41*	3.63**	4.76**	4.23**	5.03**	3.93**
STV_{vote}: ESG	3.75**	3.11**	1.34**	–0.21	1.37	5.80***	2.46	3.56	7.07**
OPT: ESG	5.76***	7.49***	8.78***	1.69	2.17	5.99***	4.58***	5.17**	4.11**

NOTES: The alphas are computed using time series regression. The standard errors are computed using heteroskedastic and autocorrelation consistent standard error estimators, with statistical significance highlighted at the 1% (***), 5% (**), and 10% (*) levels.

Among the individual-score portfolios, the MSCI portfolios consistently have the highest Fama–French five-factor alphas across all regions, while ISS portfolios have significant alphas in the United States and Japan. The aggregate ESG scores generally have higher alphas than most individual scores. These portfolios behave differently across regions and quantiles. For example, the STV_{vote} aggregate score generally has high ±10 portfolio alphas in Europe and Japan. Broadly speaking, the alphas of different aggregation methods are similar to each other, due to high correlation between their scores.

From Exhibit 7, we see there are significant excess returns (computed using the Fama–French five-factor model) of 4.8%, 2.9%, and 9.1% using individual ESG scores in the United States, Europe, and Japan, respectively, while there are excess returns of 6.1%, 6.1%, and 9.6% using aggregate ESG scores. The alpha of portfolios based on Reprisk, Truvalue Labs, and Moody's scores are not statistically significant (possibly due to the noise in the scores). However, the alphas for the aggregate ESG portfolios are significant, in general, likely due to the stronger signal from the aggregation methods.

TREYNOR–BLACK PORTFOLIOS

Along with equal-weighted portfolios, we also construct ESG portfolios using Treynor–Black weights, as given by Equation (6). In Treynor–Black portfolios, the weights are inversely proportional to the rank of the ESG score of each firm. In Exhibit 8, we include the excess returns (alpha) obtained from the time series regression using the Fama–French five-factor model and the CAPM.

Comparing the results in Exhibits 7 and 8, we do not observe a large difference between the alphas of the ±25 and ±10 portfolios, but we do find differences in alphas for the ±40 portfolios. The effect of unequal weighting becomes more prominent when more firms are included in the portfolio and when firms at extreme percentiles on the long or short sides are weighted differently. Our other observations about excess returns described in the Estimating Excess Returns subsection (in the Performance of ESG Portfolios section) remain consistent.

Combining ESG and Passive Portfolios

Once the relative weights of the securities within an ESG portfolio are determined, one can combine that portfolio with any other portfolio. For example, we can add the ESG portfolio to a suite of portfolios that mimic more traditional asset pricing factors, such as value, size, or momentum.

Perhaps the most natural application is to combine the ESG portfolio with a passive index fund such as the market portfolio. In this section, we combine the active Treynor–Black ESG portfolios with market portfolios.[15] The weight of the market portfolio is fixed to be 0.5.[16] The market portfolios we use for different regions are the MSCI USA, the MSCI Europe, and the MSCI Japan indices.

[15] Exhibit A2 in the Appendix gives the Sharpe ratio for the portfolios that are built by combining equal-weighted ESG portfolios with the passive market index.

[16] More generally, weights can also be determined by other methods. For example, Lo and Zhang (2023) show that the optimal weights to maximize the Sharpe ratio, ω, can computed by using an ESG portfolio's excess returns and idiosyncratic volatility:

$$\omega_A = \left(\frac{\alpha_A}{\sigma(\epsilon_A)^2}\right) \Big/ \left(\frac{\mathbb{E}[R_m] - R_f}{\sigma_m^2}\right), \tag{11}$$

where $\mathbb{E}[R_m]$ and σ_m^2 are the expected returns and variance of the passive portfolio, respectively.

EXHIBIT 8

Fama–French Five-Factor Model (FF5) and CAPM Alphas for Treynor–Black Top–Bottom ESG Portfolios (±40%, ±25%, ±10%)

	United States			Europe			Japan		
Portfolio	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**
FF5 Alpha									
ISS: ESG	2.62***	2.20*	2.37*	2.42	2.43	1.43	5.40**	5.94***	7.03***
MSCI: ESG	3.31**	3.92***	4.49**	2.75**	2.88**	1.74**	4.66**	5.51***	3.62***
Reprisk: ESG	1.52	1.52	3.31	–0.64	0.02	1.47	1.12	1.73	2.44**
S&P Global: ESG	0.93	0.62	1.66	2.00	2.48	3.80*	5.30***	5.95***	8.61***
TVL: ESG	1.19	0.66	–1.59	0.21	0.54	–1.28	–0.17	0.11	0.93
Moody's: ESG	2.31*	2.89*	3.94**	1.13	1.51	0.58	2.90*	2.80*	6.91**
INDI-AVG: ESG	1.98**	1.97**	2.37**	1.32	1.66	1.32	3.20**	3.67**	4.95***
AVG: ESG	4.09**	5.42***	4.64**	1.94*	2.16*	1.26*	5.87***	5.96**	6.37**
PCA: ESG	2.40*	2.88**	2.80*	2.46	2.81	3.28	5.02***	5.55**	5.90**
MAHA: ESG	4.01***	4.31**	5.17***	0.87	1.08	–0.32	5.49**	6.37***	7.62*
AVG_{vote}: ESG	3.81**	4.61***	5.12**	2.19	2.71*	2.35*	6.04***	6.63***	7.44**
STV_{vote}: ESG	1.85**	1.55**	–0.14**	1.95	3.02	6.73***	5.51**	6.52**	9.94***
OPT: ESG	4.68***	5.18***	6.19**	3.05	3.47	6.33**	6.33***	6.81***	7.08**
CAPM Alpha									
ISS: ESG	5.60***	5.37***	5.63***	3.05**	2.85**	2.93**	4.96**	5.24***	6.07**
MSCI: ESG	4.68***	5.37***	6.49***	4.34***	4.64**	4.34*	3.41**	4.10**	1.11**
Reprisk: ESG	2.04**	1.88**	3.33**	3.15	4.40*	7.34**	2.24	2.77	3.68***
S&P Global: ESG	0.45	0.24	1.92	0.39	0.73	2.63	3.22	3.67	6.10*
TVL: ESG	2.12**	1.65**	–0.98**	2.21	2.78*	1.12*	–1.12	–1.05	0.59
Moody's: ESG	2.82*	3.33**	4.47***	–0.09	0.25	–0.38	1.33	1.15	4.78
INDI-AVG: ESG	2.95***	2.98**	3.48***	2.19**	2.63**	3.03*	2.33	2.64*	3.73**
AVG: ESG	6.11**	7.79***	7.32***	3.43**	3.80**	4.10**	4.40***	4.37**	3.79**
PCA: ESG	3.67**	4.28***	4.32**	1.54	2.02	2.89	3.30	3.68	4.00
MAHA: ESG	6.10***	6.80***	7.54***	3.64*	4.27**	4.20**	4.49**	4.84**	4.82**
AVG_{vote}: ESG	5.59***	6.74***	6.85***	3.44*	4.04**	4.67**	4.44**	4.81**	4.15**
STV_{vote}: ESG	2.57**	2.19**	0.36**	2.12	3.01	6.49***	3.76	4.44	7.37**
OPT: ESG	7.09***	8.07***	9.20***	3.69	4.20	7.22***	5.04***	5.32**	4.46**

NOTES: The alphas are computed using time series regression. The standard errors are computed using heteroskedastic and autocorrelation consistent standard error estimators, with statistical significance highlighted at the 1% (***), 5% (**), and 10% (*) levels.

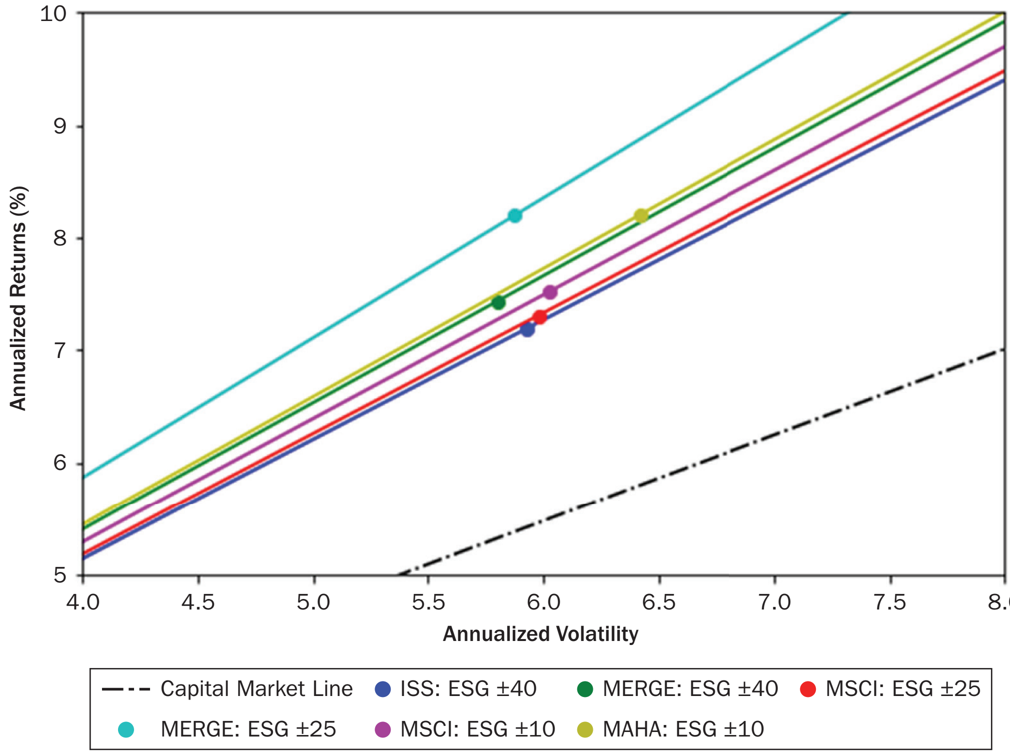
Exhibit 9 presents the expected returns and volatility of several combined ESG portfolios. In each region, we include the top–bottom portfolios (±40, ±25, and ±10) with the highest Sharpe ratio, based on single or aggregate ESG scores. The Sharpe ratios of the combined portfolios are higher than those of market portfolios due to the signal in the ESG scores. These improved Sharpe ratios are not accessible to traditional mean–variance optimized portfolios, which stay below the capital market line. This forms a "super-efficient frontier" compared with the capital market line associated with the passive portfolio, assuming that the alphas from the ESG portfolios are mispricings. Under the alternate interpretation that ESG scores capture an omitted pricing factor, the "super-efficiency" of the new frontier may be viewed as the result of additional risk premia not accessible to investors except for ESG portfolio managers.

In particular, the combined portfolios achieve annual Sharpe ratios as high as 1.25, 0.53, and 0.72 in the United States, Europe, and Japan, respectively. As a comparison, the Sharpe ratios of the market portfolios are only 0.75, 0.05, and 0.37. Across all portfolios, the Sharpe ratios are always positive in the United States and Japan, while in Europe, they are negative for portfolios based on Moody's scores.

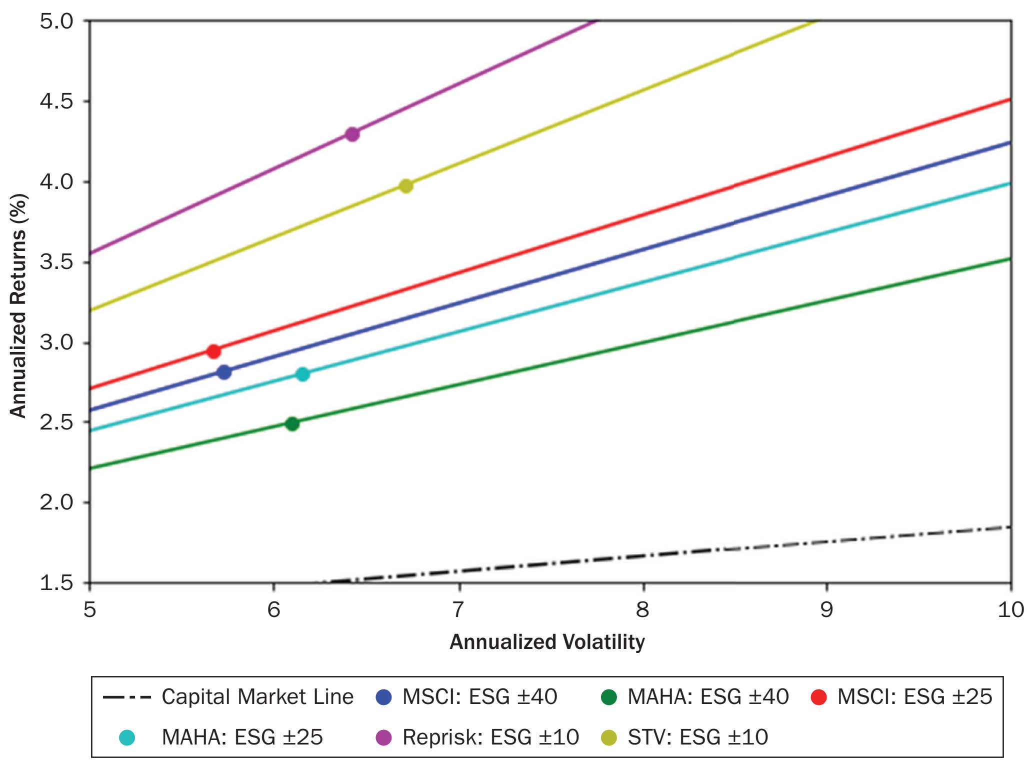
EXHIBIT 9

Annualized Returns (y-axis) versus Volatility (x-axis)

Panel A: United States



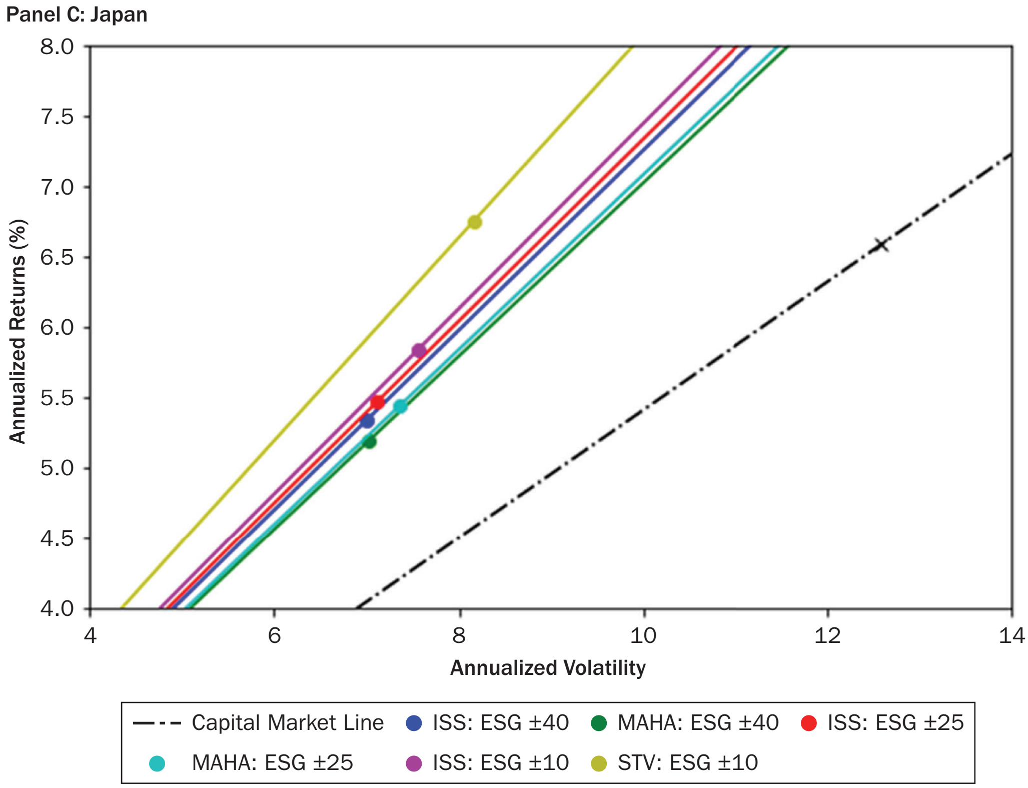
Panel B: Europe



(continued)

EXHIBIT 9 *(continued)*

Annualized Returns (y-axis) versus Volatility (x-axis)



NOTES: We present the returns versus volatility for portfolios that combine a Treynor–Black ESG portfolio with the market portfolio. The capital market line is represented using the black dash-dotted line. The combined portfolios lie above the capital market line. The entries on the legend are in the format: "[ESG Dataset] [Portfolio Name]"

ESG AS UNIVARIATE IMPACT FACTORS

We now turn to analyzing portfolios constructed using univariate scores. Like our analysis of portfolios based on full ESG scores, we construct top–bottom sorted portfolios using the E, S, and G scores individually, and compute their excess returns.

Environmental Portfolios

We include the excess returns for environmental portfolios in Exhibit 10. For environmental scores, we observe high and statistically significant alphas (up to 10.65%) for portfolios based on individual scores and for portfolios based on aggregate scores in Japan. For the United States, individual scores do not achieve significant alpha, while aggregate scores generate significant positive alpha, with excess returns of up to 4.5%. For Europe, we do not observe significant alphas (except for a few portfolios). CAPM alphas are positive, significant, and higher for the United States compared with other regions.

Social Portfolios

We find there are similar patterns for portfolios based on social scores (see Exhibit 11). However, these portfolios have negative significant excess returns (CAPM alpha) in Europe. Fama–French five-factor annualized alphas for S-score

EXHIBIT 10

Fama–French Five-Factor Model (FF5) and CAPM Alphas for Top–Bottom Environmental (E) Portfolios (±40%, ±25%, ±10%)

	United States			Europe			Japan		
Portfolio	±40%	±25%	±10%	±40%	±25%	±10%	±40%	±25%	±10%
FF5 Alpha									
ISS: E	2.46*	0.76*	–0.17*	1.79	1.13	–0.04	3.21	4.19*	8.21***
MSCI: E	3.12*	4.24*	8.72**	0.73	0.24	0.16	3.68***	4.55***	6.54***
Reprisk: E	0.55	2.22	0.45	–1.70	–2.65*	–0.90*	3.06*	2.53*	1.41*
S&P Global: E	1.60	1.66	2.20	0.01	1.36	4.02**	4.34***	5.82***	10.45***
Moody's: E	0.74	1.39	–2.35	2.05**	1.26**	–0.58**	2.11	1.83	4.42
AVG: E	2.98**	3.95**	4.49**	0.63	0.42	0.84	3.45*	7.30***	9.99***
PCA: E	2.50**	1.78**	–2.10**	0.78	1.37	0.48	3.71***	4.31***	7.54***
MAHA: E	1.67	3.30*	5.83*	1.37	1.16	0.50	4.69**	7.37***	9.65***
AVG_{vote}: E	2.29*	3.34*	4.16*	1.09	1.61	1.63	3.40*	6.94***	8.88***
STV_{vote}: E	1.95	1.88	2.26	0.06	0.33	3.53*	5.15***	6.74***	10.65***
CAPM Alpha									
ISS: E	4.96***	3.37***	2.93***	2.44**	2.15**	2.30**	2.97	3.99*	7.24**
MSCI: E	5.63**	7.72**	13.68***	0.75	0.51	0.97	2.55	3.39*	5.30*
Reprisk: E	1.66	3.17*	1.92*	0.63	0.32	4.27	4.74***	4.19**	4.09**
S&P Global: E	2.69*	2.50*	3.85**	–1.26	–0.30	2.67*	1.90	3.34*	7.87**
Moody's: E	1.94**	3.12***	–1.31***	1.05	–0.04	–1.03	0.57	0.15	3.93*
AVG: E	5.25***	7.50***	10.27***	1.39	0.78	2.13	2.62*	5.86***	8.42**
PCA: E	4.22***	3.82**	0.14**	–0.27	0.14	–0.55	1.88	2.33	6.03**
MAHA: E	3.67***	6.32***	10.96***	2.62	3.26	3.33	4.50***	6.86***	8.95***
AVG_{vote}: E	4.47***	6.61***	10.05***	1.27	1.63	2.73	2.17	5.40**	7.66**
STV_{vote}: E	3.19*	3.13*	5.51***	–1.03	–0.88	2.70	3.26**	4.61**	7.66**

NOTES: The alphas are computed using time series regression. The standard errors were computed using heteroskedastic autocorrelation consistent standard error estimators, with statistical significance highlighted at the 1% (***), 5% (**), and 10% (*) levels.

portfolios are positive and significant in the United States (up to 4.3%) and Japan (up to 7.6%).

Governance Portfolios

For governance-score portfolios (see Exhibit 12), the aggregation does not produce significant Fama–French five-factor alphas in the United States or Europe compared with individual vendor scores; however, we find Fama–French five-factor excess returns of up to 11.8% in Japan.

CONCLUSION

Using the framework proposed by Lo and Zhang (2023), we quantify the excess returns of arbitrary ESG portfolios via the cross-sectional standard deviation of the stock's excess returns and the correlation between the excess returns and ESG factors (both combined and individual E, S, and G scores) obtained from six leading ESG score providers for firms in the United States, Europe, and Japan from 2014 to 2020. Few studies have analyzed such a comprehensive dataset and as systematically. We also propose a number of methods to aggregate ESG scores across vendors to

EXHIBIT 11

Fama–French Five-Factor Model (FF5) and CAPM Alphas for Top–Bottom Social (S) Portfolios (±40%, ±25%, ±10%)

	United States			Europe			Japan		
Portfolio	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**
FF5 Alpha									
MSCI: S	0.86	0.37	3.04**	1.70	1.95	–0.62	2.82**	2.66**	2.22**
Reprisk: S	1.76	1.84	0.05	–0.72	–0.14	1.20	0.43	–0.60	–1.28
S&P Global: S	1.42	1.58	–3.40	–0.78	0.15	–1.61	3.87**	5.12***	7.61***
Moody's: S	2.56**	3.28**	1.67**	–0.67	–0.80	1.60	3.44**	0.94**	3.13**
AVG: S	2.69***	3.64***	3.83*	0.44	0.11	–2.32	2.64**	3.81*	7.10**
PCA: S	1.81	1.51	–0.30	–0.88	0.20	0.87	2.79**	3.89***	5.10**
MAHA: S	3.93***	4.28***	2.54*	0.64	1.13	–1.95	3.00**	3.01*	6.19*
AVG_{vote}: S	2.29***	2.51**	4.30***	0.98	–0.48	–2.32	3.45**	4.97**	5.81*
STV_{vote}: S	1.31	1.49	–1.37	–0.43	0.31	–1.36	3.64**	5.68***	7.48***
CAPM Alpha									
MSCI: S	0.90	–0.40	1.49	1.11	1.58	–0.80	1.19	0.57	0.05
Reprisk: S	2.35**	2.13**	–1.29**	1.71	2.98*	5.41*	1.96	1.53	1.21
S&P Global: S	1.46	1.43	–3.20	–2.34***	–2.01*	–3.23*	1.43	3.15	5.13
Moody's: S	3.43**	4.23**	2.71**	–2.59***	–2.62**	–1.14**	1.18	–1.03	1.51
AVG: S	3.49***	4.50***	3.98*	–0.40	–0.82	–1.56	1.23	2.11	4.73
PCA: S	2.49*	1.67*	–0.06*	–3.03***	–2.50***	–2.84***	0.34	1.02	1.95
MAHA: S	5.19***	5.97***	3.29*	0.78	1.77	–0.14	1.95	2.62	4.60
AVG_{vote}: S	2.88***	2.71**	4.28**	–0.37	–1.62	–2.25	1.79	2.57	3.20
STV_{vote}: S	1.37	1.67	–0.47	–1.93***	–1.76***	–2.72***	1.31	3.61	6.00

NOTES: The alphas are computed using time series regression. The standard errors were computed using heteroskedastic autocorrelation consistent standard error estimators, with statistical significance highlighted at the 1% (***), 5% (**), and 10% (*) levels.

produce the best signal within the data, simultaneously addressing measurement errors and yielding a single measure of ESG that can potentially be used for portfolio management.

Empirically, we find significant ESG excess returns in the United States and Japan. We also find positive and higher-than-market risk-adjusted returns. We construct an aggregate ESG measure based on a linear combination of ESG scores that is optimized to maximize the correlation with excess returns. The ESG portfolio properties of the optimized ESG score are comparable to the aggregate scores, implying that our methods of aggregation were successful in amplifying the signal. We evaluate the properties of ESG portfolios by investigating their exposure to various risk factors, constructing optimal Treynor–Black-weighted portfolios, and combining them optimally with passive index portfolios, which yields "super-efficient" frontiers that all investors should be interested in accessing.

One practical implication from our results is that aggregation methods help to reduce the noise and amplify the signal contained in E, S, and G metrics to yield better estimates of ESG portfolio properties, even though individual ESG ratings are noisy and the portfolios constructed using ESG scores from any single vendor may also be quite noisy. This aggregation method can be selected at the preference of the portfolio manager because different methods will weight the noise and signal from rating agency scores in different ways. Because the true noise and signal component remain unknown, however, it is hard to establish the superiority of any particular aggregation method, and we leave this important topic for future research.

EXHIBIT 12

Fama–French Five-Factor Model (FF5) and CAPM Alphas for Top–Bottom Governance (G) Portfolios (±40%, ±25%, ±10%)

	United States			Europe			Japan		
Portfolio	±40%	±25%	±10%	±40%	±25%	±10%	±40%	±25%	±10%
FF5 Alpha									
ISS: G	1.89**	2.32*	3.98*	–0.37	0.31	–0.19	4.18***	4.12***	3.91**
MSCI: G	2.43**	3.66***	4.01***	1.08	0.17	1.90	1.99	2.99**	5.23***
Reprisk: G	0.34	2.32**	1.79*	0.51	0.71	4.60*	–0.14	1.34	4.89***
S&P Global: G	–0.25	0.42	2.72	–0.13	1.00	3.98**	3.45**	5.12***	11.75***
Moody's: G	0.29	1.21	1.30	0.83	0.45	0.83	1.71	2.88	3.09
AVG: G	2.25**	3.44**	5.84**	0.75	0.81	0.31	4.44***	5.01***	6.68*
PCA: G	2.05	1.39	0.86	0.26	1.28	2.79	0.18	–0.44	–3.86
MAHA: G	1.57	1.91	4.27*	0.83	0.54	–1.33	3.33*	5.44***	3.45***
AVG_{vote}: G	1.74*	2.43*	3.72*	0.79	0.85	0.31	4.50***	4.85**	7.42***
STV_{vote}: G	–0.28	0.40	2.58	0.18	1.36	3.98**	3.46**	5.05***	9.79***
CAPM Alpha									
ISS: G	2.32**	2.65**	4.02**	1.15	1.79	1.20	3.60***	3.24***	2.40***
MSCI: G	1.51	2.51**	1.91**	3.09***	3.34**	7.46**	1.57	2.78	5.31***
Reprisk: G	0.42	2.22**	1.19**	2.61	3.36*	8.61***	0.85	2.89	6.78***
S&P Global: G	–2.26	–0.82	3.30	–1.86*	–1.69**	0.29**	1.15	1.80	7.56***
Moody's: G	–0.08	0.60	–0.43	0.67	–0.13	1.36	–0.29	0.61	–1.06
AVG: G	1.42	2.19	4.03	2.64***	3.51*	2.45*	3.39**	3.36*	5.58*
PCA: G	1.49	0.58	0.19	–0.27	1.03	2.00***	–0.33	–0.48	–3.98*
MAHA: G	1.01	0.69	1.95	2.71**	3.33**	1.48**	2.77*	4.54**	2.62**
AVG_{vote}: G	0.60	1.20	2.06	2.65***	3.38**	1.74**	3.25**	2.98*	4.70*
STV_{vote}: G	–2.22	–0.55	2.50	–1.36	–1.22	0.87	1.29	2.02	5.82**

NOTES: The alphas are computed using time series regression. The standard errors were computed using heteroskedastic autocorrelation consistent standard error estimators, with statistical significance highlighted at the 1% (***), 5% (**), and 10% (*) levels.

Perhaps the most important practical implication is the use of our estimates to provide proper disclosure to investors as to the specific performance impact that ESG investing could have on their wealth. The recent controversy surrounding ESG funds—and the accusations of potential violations of fiduciary duty that ESG critics have leveled against ESG managers—can be addressed in part by disclosing the estimated alpha associated with various ESG constraints. By allowing investors to vote with their dollars for (or against) specific types of impact after disclosing the financial consequences that such impact implies, managers can do well by investors while doing good for society, a fitting tribute to the legacy of Harry M. Markowitz.

APPENDIX

RANK AUTOCORRELATION

In Exhibit A1, we present the cross-sectional rank autocorrelation between vendor ESG scores and aggregate scores averaged over time. We find there is a 99% autocorrelation computed with a delay of one month, implying that ESG scores do not change significantly over short periods. However, lower values of rank autocorrelation at longer delay windows implies that scores change significantly over longer time windows. This pattern is consistent across regions. Hence, to rebalance our ESG portfolios, we use a time window of 12 months.

EXHIBIT A1

Cross-Sectional Rank Autocorrelation between Vendor ESG Scores and Aggregate Scores

	United States				Europe				Japan			
Delay (Months):	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**	**1**	**3**	**6**	**12**
ISS: ESG	99	98	97	94	99	98	97	95	99	98	97	95
MSCI: ESG	99	97	95	90	99	97	94	90	99	97	94	89
Reprisk: ESG	98	96	93	88	98	96	93	86	98	94	89	79
S&P Global: ESG	99	98	96	92	99	98	97	94	99	99	98	96
TVL: ESG	98	94	86	73	99	95	88	75	99	94	85	69
Moody's: ESG	99	98	96	94	99	98	97	95	99	98	96	94
AVG: ESG	99	98	96	92	99	98	97	94	99	98	96	92
PCA: ESG	99	99	98	97	99	99	98	97	99	99	98	97
MAHA: ESG	98	96	93	86	98	96	92	86	98	95	91	82
AVG_{vote}: ESG	99	98	96	93	99	98	97	94	99	98	96	92
STV_{vote}: ESG	95	88	82	71	95	90	83	74	96	91	85	76

EXHIBIT A2

Sharpe Ratios of Combined Market and ESG Treynor–Black Portfolios

	United States			Europe			Japan		
Portfolio	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**	**±40%**	**±25%**	**±10%**
Market:		**0.75**			**0.05**			**0.37**	
ISS: ESG	1.07	1.03	1.01	0.20	0.18	0.18	**0.64**	**0.65**	0.66
MSCI: ESG	1.03	1.08	1.10	**0.33**	**0.36**	0.32	0.54	0.58	0.34
Reprisk: ESG	0.81	0.79	0.87	0.22	0.31	**0.53**	0.45	0.48	0.52
S&P Global: ESG	0.68	0.66	0.76	0.01	0.03	0.16	0.49	0.51	0.62
TVL: ESG	0.80	0.75	0.54	0.14	0.18	0.06	0.23	0.24	0.33
Moody's: ESG	0.87	0.91	0.95	–0.03	–0.01	–0.05	0.38	0.36	0.55
INDI-AVG: ESG	0.91	0.91	0.94	0.14	0.18	0.21	0.47	0.49	0.56
AVG: ESG	**1.13**	**1.25**	1.13	0.24	0.26	0.28	0.59	0.58	0.50
PCA: ESG	0.94	0.98	0.94	0.09	0.12	0.18	0.51	0.52	0.51
MAHA: ESG	1.12	1.15	**1.14**	0.26	0.31	0.30	0.62	0.62	0.57
AVG_{vote}: ESG	1.09	1.16	1.10	0.24	0.28	0.33	0.58	0.59	0.51
STV_{vote}: ESG	0.84	0.79	0.63	0.14	0.20	0.46	0.55	0.57	**0.72**
OPT: ESG	1.21	1.26	1.25	0.25	0.29	0.52	0.64	0.65	0.53

NOTES: The market portfolios used for different regions are the MSCI USA, MSCI Europe, and MSCI Japan indices. The weight of the market index is fixed at 0.5. The highest realizable Sharpe ratio in each portfolio is shown in bold font.

Exhibit A2 gives the Sharpe ratio for the portfolios that are built by combining equal-weighted ESG portfolios with the passive market index.

ACKNOWLEDGMENTS

We thank Jayna Cummings for editorial assistance and seminar and conference participants at the 2023 MIT Sloan ESG & Impact Finance Conference, the finance workshop at Peking University's Guanghua School of Management, and the MIT Sloan Sustainability Research Lunch. Ruixun Zhang acknowledges research funding from the National Key R&D Program of China (2022YFA1007900) and the National Natural Science Foundation of China (12271013,72342004). The views and opinions expressed in this article are those of the authors only and do not necessarily represent the views and opinions of any other organizations, any of their affiliates or employees, or any of the individuals acknowledged above. We thank all anonymous reviewers for their insightful comments and suggestions, and are grateful to Massachusetts Pension Reserves Investment Management Board, AQR Capital Management (Applied Quantitative Research), SimCorp, Asset One, and MFS Investment Management—members of the Aggregate Confusion Project Council—for their generous support of this research. We are also extremely grateful to the ESG rating agencies that generously provided their data to this project.

REFERENCES

Albuquerque, R., Y. Koskinen, and C. Zhang. 2019. "Corporate Social Responsibility and Firm Risk: Theory and Empirical Evidence." *Management Science* 65 (10): 4451–4469.

Berg, F., K. Fabisik, and Z. Sautner. 2021. "Is History Repeating Itself? The (Un) Predictable Past of ESG Ratings." ECGI Finance Working Paper No. 708/2020 (August).

Berg, F., F. Heeb, and J. F. Kölbel. 2022. "The Economic Impact of ESG Rating Changes." Working paper, SSRN 4088545.

Berg, F., J. F. Kölbel, A. Pavlova, and R. Rigobon. 2021. "ESG Confusion and Stock Returns: Tackling the Problem of Noise." NBER Working Paper Series No. 30562.

Berg, F., J. F. Kölbel, and R. Rigobon. 2022. "Aggregate Confusion: The Divergence of ESG Ratings." *The Review of Finance* 26 (6): 1315–1344.

Bhattacharya, P. 1974. "Convergence of Sample Paths of Normalized Sums of Induced Order Statistics." *Annals of Statistics* 2 (5): 1034–1039.

Bolton, P., and M. T. Kacperczyk. 2020. "Carbon Premium around the World." CEPR Discussion Paper No. DP14567.

Brandon, R. G., P. Krueger, and P. S. Schmidt. 2021. "ESG Rating Disagreement and Stock Returns." *Financial Analysts Journal* 77 (4): 104–127.

Chava, S. 2014. "Environmental Externalities and Cost of Capital." *Management Science* 60 (9): 2223–2247.

Dorfleitner, G., G. Halbritter, and M. Nguyen 2015. "Measuring the Level and Risk of Corporate Responsibility—An Empirical Comparison of Different ESG Rating Approaches." *Journal of Asset Management* 16 (7): 450–466.

The Economist. "ESG Investing: A Broken Idea." Special Report (July 23, 2022): https://www.economist.com/special-report/2022-07-23.

Edmans, A. 2011. "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices." *Journal of Financial Economics* 101 (3): 621–640.

El Ghoul, S., O. Guedhami, C. C. Kwok, and D. R. Mishra. 2011. "Does Corporate Social Responsibility Affect the Cost of Capital?" *Journal of Banking & Finance* 35 (9): 2388–2406.

Khan, M., G. Serafeim, and A. Yoon. 2016. "Corporate Sustainability: First Evidence on Materiality." *Accounting Review* 91 (6): 1697–1724.

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." *The Journal of Finance* 72 (4): 1785–1824.

Lo, A. W., and A. C. MacKinlay. 1990. "Data-Snooping Biases in Tests of Financial Asset Pricing Models." *The Review of Financial Studies* 3 (3): 431–467.

Lo, A. W., L. Wu, R. Zhang, and C. Zhao. 2024. "Optimal Impact Portfolios with General Dependence and Marginals." *Operations Research* (ahead of print). https://doi.org/10.1287/opre.2023.0400.

Lo, A. W., and R. Zhang. 2023. "Quantifying the Impact of Impact Investing." *Management Science* (ahead of print). https://doi.org/10.1287/mnsc.2022.01168.

Pástor, L., and R. F. Stambaugh. 1999. "Costs of Equity Capital and Model Mispricing." *The Journal of Finance* 54 (1): 67–121.

Pástor, L., R. F. Stambaugh, and L. A. Taylor. 2021. "Sustainable Investing in Equilibrium." *Journal of Financial Economics* 142 (2): 550–571.

——. 2022. "Dissecting Green Returns." *Journal of Financial Economics* 146 (2): 403–424.

Semenova, N., and L. G. Hassel. 2015. "On the Validity of Environmental Performance Metrics." *Journal of Business Ethics* 132 (2): 249–258.