

## Secondary Publication



Horn, Matthias; Oehler, Andreas

### Constructing stock portfolios by sorting on ESG ratings : Does the rating provider matter?

Date of secondary publication: 19.03.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1070760

#### Primary publication

Horn, Matthias; Oehler, Andreas (2024): Constructing stock portfolios by sorting on ESG ratings : Does the rating provider matter?, in: International review of financial analysis, Amsterdam [u.a.]: Elsevier, Vol. 96, Part A, Nr. 103568, pp. 1–30, doi: 10.1016/j.irfa.2024.103568.

#### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>



# Constructing stock portfolios by sorting on ESG ratings: Does the rating provider matter?<sup>☆</sup>

Matthias Horn<sup>a,\*</sup>, Andreas Oehler<sup>b</sup>

<sup>a</sup> Department of Finance, Bamberg University, Bamberg, Germany

<sup>b</sup> Bamberg University, Bamberg, Germany

## ARTICLE INFO

### JEL codes:

G11

G24

### Keywords:

ESG rating

Rating dispersion

Portfolio performance

Idiosyncratic risk

Sustainability

Best-in-class

## ABSTRACT

An interesting and well documented observation is that the same company receives heterogeneous ESG ratings from different rating providers. Consequently, an important question is whether sorting stocks on ESG ratings of different rating providers results in portfolios that are similar or substantially different regarding their constituents, performance, and risk. We employ the ratings of five rating providers on stocks listed in North America, Europe, Asia-Pacific (excluding Japan), and Japan for the period from 2014 until the end of 2019 and analyze whether the corresponding quintile stock portfolios actually have similar or different portfolio constituents as well as Sharpe ratios, alphas, and idiosyncratic risk. Our analysis indicates that the portfolios considerably differ regarding their constituents. In contrast, and most importantly, Sharpe ratios, alphas, and idiosyncratic risk of corresponding portfolios are not significantly different.

## 1. Introduction

Recent evidence suggests that sentiment toward ESG investing is reversing in some countries. Early studies indicate that assets with high ESG ratings had attracted significant investment inflows,<sup>1</sup> while more recent analyses reveal that investors are avoiding or withdrawing money from ESG funds (Temple-West & Schmitt, 2024). Some of this recent aversion to ESG investments can be explained with a shift in investment behavior. Pension and mutual fund manager in the U.S. have come under immense political pressure to focus solely on financial performance excluding other considerations. Consequently, BlackRock, Vanguard and other major asset managers nowadays shy away from promoting ESG as they fear a decline in assets under management (AuM). Moreover, there is some theoretical and empirical evidence that ESG funds should underperform in normal times (Engle et al., 2020) questioning the benefits of ESG investments for the private investor in normal times. Finally, the Russian invasion of Ukraine resulted in high returns for the defense industry as well as for fossil energy companies, which may have led investors to reconsider their optimal asset

allocation strategies. Overall, it remains an open issue whether ESG investment strategies are effective in achieving a long-term financial outperformance in normal times.

Most importantly, successful ESG investment strategies have to rely on precise ESG ratings, intended to assess a company's ESG performance or risk. These ratings are often considered unreliable due to data imputation, revisions of historical data, and their typical annual updates (Berg et al., 2020; Kotsantonis & Serafeim, 2019). Even in weakly efficient financial markets, such infrequent, publicly available data would be deemed an insufficient basis for a successful investment strategy (Fama, 1991).

Despite all these obvious problems, a significant proportion of investors still believe that stocks with high ESG ratings will outperform the market in the long run (Giglio, Maggiori, et al., 2023). This discrepancy between financial theory and investor beliefs and behavior has prompted numerous studies to analyze whether incorporating ESG ratings into portfolio optimization strategies affects the likelihood of achieving persistent outperformance. The results of these studies are mixed at best and do not provide conclusive evidence that the relationship between

<sup>☆</sup> We would like to thank Brian Lucey and Samuel Vigne (the editors), Wolfgang Bessler (the guest editor), two anonymous referees, Stefan Wendt, Julian Schneider, and participants of the New Frontiers in Banking and Capital Markets Conference 2022 in Rome for helpful comments and suggestions, and Thomas Walker for helpful comments and suggestions and support with the data. All remaining errors are our own.

\* Corresponding author at: Department of Finance, Bamberg University, Kaerntenstrasse 7, 96045 Bamberg, Germany.

E-mail address: [matthias.horn@uni-bamberg.de](mailto:matthias.horn@uni-bamberg.de) (M. Horn).

<sup>1</sup> Benson and Humphrey (2008); Bialkowski and Starks (2016); Hartzmark and Sussman (2019); Latino et al. (2021).

ESG ratings and risk-adjusted returns is positive, negative, or only temporarily significant. Therefore, further research is required to provide a clearer picture for investors.

In this study, we focus on one aspect of ESG ratings that has only recently garnered attention: the heterogeneity of ESG ratings offered by different providers, each employing distinct methodologies or rating approaches. Consequently, ESG ratings for a single company can differ across providers.<sup>2</sup> Our research is motivated by the need for understanding the portfolio and performance effects of ESG rating dispersion. For this we investigate the financial outcomes for investors when ESG ratings are utilized in portfolio optimization. Specifically, we examine the common technique to sort stocks by ESG ratings when ESG is considered as an important element in portfolio construction. The idea behind our approach to analyze the differences between high and low ESG firms is that some investors prefer to invest only in the stocks with the highest ESG ratings, i.e. the most sustainable stocks.

Additionally, as some investors focus on investing in the most sustainable ESG funds, other investors may seek to invest specifically in the least sustainable stocks, those with low ESG ratings, envisioning some outperformance. We address whether stock selection based on ESG ratings can enhance financial performance. Specifically, the primary contribution of our paper is to compare the ESG ratings from different providers, analyzing whether the ratings from one provider are superior when used for portfolio optimization.

Given the efficiency of financial markets and the criticism regarding the quality of ESG ratings, we hypothesize that long-short portfolios, consisting of long positions in high-ESG stocks and short positions in low-ESG stocks, do not, on average, outperform the market. Similarly, we further hypothesize that no single ESG rating is superior to others for portfolio construction; that is, portfolios based on the ESG ratings from one provider do not outperform those based on the ESG ratings from other providers.

One important issue that is less researched is that the relationship between ESG ratings and portfolio performance may vary across geographic regions. Surveys indicate that investors' ESG preferences differ significantly across regions and countries (Starks, 2023). Additionally, country characteristics are a primary driver of differences in companies' ESG ratings, as cultural norms influence the behavior of investors, corporate executives, and policymakers.<sup>3</sup> External shocks that lead to financial benefits for stocks of companies with high ESG ratings should therefore have varied effects on stocks in different countries. Therefore, we extend our previous hypotheses on the usefulness of ESG ratings for improving financial performance and the potential superiority of one ESG rating provider by investigating whether these results dependent on the region in the world. We hypothesize that long-short portfolios based on ESG ratings will outperform the market in some regions, while underperforming the market in others, resulting in an overall lack of outperformance. If we find support for this hypothesis, this could explain the mixed evidence on the relationship between ESG ratings and risk-adjusted returns. Furthermore, we hypothesize that portfolios based on the ESG ratings of one provider do not outperform those based on the ratings of other providers in any region.

To test these hypotheses, we employ a dataset of seven stock market indices from various regions and countries and construct stock portfolios based on the ESG ratings of five popular rating providers: ISS Oekom, MSCI KLD, Sustainalytics, Refinitiv, and Vigeo Eiris. In our analyses, we consider three different investor perspectives: first, an investor implementing a naïve diversification strategy by equally weighting the stocks in the portfolio; second, an investor following a market-oriented strategy by weighting stocks according to their market value; and third, an investor adopting a risk-mitigating approach using a minimum-variance

optimization model. We compare the respective portfolios in terms of the intersection of their constituents, returns, Sharpe ratios, alphas, and idiosyncratic risk.

This study contributes to the literature by comparing the financial performance of portfolios constructed using ESG ratings from different providers across various global regions. The analysis spans from January 2014 to December 2019. Our initial analysis reveals significant differences in the coverage of ESG rating providers. One approach would be to include only stocks covered by all five rating agencies. However, this would substantially reduce the sample size and limit its informativeness and conclusions. To address this problem, we employ a pairwise approach in our analyses, comparing the outcomes of two ESG ratings at a time (e.g., ISS with MSCI, ISS with Sustainalytics, etc.), and including only stocks that have received ratings from both agencies. This pairwise approach offers an additional advantage and further contributes to the literature by ensuring that differences in portfolio constituents or performance are not due to varying coverage by the rating providers, as we only consider the intersection of stocks rated by both providers. As the available portfolio constituents also change over time due to adjustments in the coverage by ESG rating providers we account for these changes and potential selection bias by also constructing portfolios of stocks with only one or no ESG rating.

For all four geographic regions, we find that, *ceteris paribus*, sorting based on different ESG ratings results in portfolios with remarkably different constituents. Our results reveal that two portfolios consisting of the top ESG quintile (the 20 % of stocks with the highest ESG ratings) from two rating providers overlap by only 52 %. This means that only 52 % of the stocks rated as the most sustainable by one provider are also rated as the most sustainable by another provider. The alphas of equally weighted long-short portfolios are, on average, negative in the Asia-Pacific region, slightly negative in North America, zero in Europe, and positive in Japan. The average alpha of the equally weighted long-short portfolios across all regions is slightly negative. The alphas of market-cap-weighted long-short portfolios are negative in the Asia-Pacific and North America, and zero in Europe and Japan. Minimum variance optimized long-short portfolios show negative alphas. These results support the hypothesis that long-short portfolios based on ESG ratings outperform in some markets and underperform in others, leading to an overall lack of outperformance for long-short portfolios on average.

Although portfolios within the same rating quintile and region differ substantially in terms of their constituents, they do not differ significantly in terms of return, alpha, Sharpe ratio, and idiosyncratic risk. This result is consistent regardless of whether the portfolios are sorted based on ISS, MSCI KLD, Refinitiv, Sustainalytics, or Vigeo Eiris ESG ratings. Therefore, these findings strongly support the hypothesis that portfolios based on ESG ratings from one provider do not outperform portfolios based on ESG ratings from other providers in any region.

Our results suggest that if an investor must choose an ESG rating provider and aims to achieve superior risk-adjusted returns, the universe of rated stocks is much more important than the rating approach that determines the overall ESG rating. Even if we find some portfolios that outperform, it is important to note that maximizing performance by investing in stocks with a single, no, or low ESG rating likely involves exposure to long-term risks such as physical climate risk or transition risk. If investors want to hedge against climate risk, they will likely have to invest in a portfolio that underperforms in the absence of severe climate news or regulation. However, when climate risks materialize, a portfolio of stocks with high ESG ratings from different providers is expected to outperform (De Nard et al., 2024; Engle et al., 2020).

The remainder of our paper proceeds as follows: Section 2 reviews the related literature, and Section 3 briefly summarizes studies on ESG ratings and describes the employed ESG rating data. In Section 4, we describe the stock data and methodology and report descriptive statistics. The empirical findings are presented and discussed in Section 5. Section 6 concludes.

<sup>2</sup> Dorfleitner et al. (2015); Chatterji et al. (2016); Kotsantonis and Serafeim (2019); Billio et al. (2021); Berg et al. (2021, 2022); Oehler and Horn (2021).

<sup>3</sup> Cai et al. (2016); Dyck et al. (2019); Graham (2022).

## 2. Literature review

A significant proportion of professional and retail investors indicate a preference for sustainable stocks. These are the stocks of those companies with good environmental performance (E), a commitment to social responsibility (S), and good corporate governance (G) (Bauer et al., 2020; Brodback et al., 2019). These preferences for sustainable stocks may be driven by personal values (nonpecuniary preferences) and/or financial expectations (pecuniary preferences) (Starks, 2023).

Investors driven by personal values either want companies to transform their business models into environmentally, ethically, and socially responsible models or to divest from unethical business models and invest only in high-ESG stocks. In our empirical analysis, we do not consider nonpecuniary preferences, such as impact or altruistic motives, for the following reasons. Nonpecuniary preferences vary widely among investors and are related to several personal characteristics (Assaf et al., 2024; Hervé & Marsat, 2024). This variability makes it difficult to determine whether an investment has met investors' expectations, even *ex post*. Moreover, relatively low passing rates for shareholder proxy proposals on ESG issues indicate that investors need at least one large, influential institutional investor, as studied by Hoepner et al. (2024), to impact firms' ESG policies. However, there are no indications that even mutual funds with an ESG mandate would take on an active role and force companies to become more sustainable (Raghunandan & Rajgopal, 2022). Instead, ESG funds tend to divest from firms following an ESG rating downgrade rather than pressuring the firm to take action to obtain an upgrade. Firms seek to improve their sustainability performance only when doing so would increase shareholder value (Berg et al., 2024; Clementino & Perkins, 2021). This could be the case as most investors are unwilling to pay significantly more for greater impact (Heeb et al., 2023). Therefore, we focus exclusively on the financial performance of ESG investment approaches.

Previous studies analyze the financial performance of ESG investment approaches using three different frameworks: asset pricing models in market equilibrium, hedge portfolios, and portfolio optimization models. These approaches are not strictly distinct from each other. Specifically, the asset pricing models by Pástor et al. (2021), Pedersen, Fitzgibbons, and Pomorski (2021), and Zerbib (2022) provide general insights that help explain the financial outcomes of hedge portfolios and portfolio optimization models.

These models incorporate two key aspects. First, due to investors' nonpecuniary preferences and social norms, some investors abstain from sin stocks or low-ESG stocks (Hong & Kacperczyk, 2009). Second, high-ESG stocks face lower climate risk and lower litigation risk, heightened by social norms. According to Merton's (1987) equilibrium model, these two aspects should cause low-ESG stocks to trade at a discount and yield higher returns in the future. The mentioned studies find empirical evidence supporting this pattern. However, it is important to note that high-ESG stocks outperform when investors' tastes for high-ESG stocks increase or when ESG risks materialize (Ardia et al., 2023; Pástor et al., 2022). The former pattern indicates that investors solely concerned with financial performance can optimize their portfolios by investing in the least sustainable stocks and by anticipating shifts in investor preferences toward high-ESG stocks. The latter pattern indicates that high-ESG stocks can serve as a hedge against ESG risk.

As investor awareness of ESG risk has increased, demand for hedging these risks has soared (Giglio et al., 2021; Krueger et al., 2020; Stroebel & Wurgler, 2021). However, instruments to directly hedge some ESG risks are illiquid and only available to institutional investors. Consequently, many investors' only option to manage ESG risks is through hedge portfolios (Alekseev et al., 2022; Engle et al., 2020). Hedge portfolios are considered an insurance-like risk management tool. Some ESG risks, such as climate risk, may materialize at unknown times with uncertain effect sizes. Additionally, the timing and nature of policies to mitigate these risks are uncertain and likely to negatively affect the financial performance of firms.

Investors can focus on individual ESG risk components or on ESG risks in general. When focusing on individual ESG risk components, such as climate risk, investors can create a hedge portfolio that mimics the respective risk (De Nard et al., 2024). However, ESG risks are diverse and include carbon risk, climate risk, biodiversity risk, water risk, etc. (Andersson et al., 2016; Giglio, Kuchler, et al., 2023). Given the uncertainty of investors' exposure to individual ESG risk components, we focus on a general ESG risk hedge. This involves divesting from low-ESG stocks and holding a portfolio of high-ESG stocks as a hedge portfolio.

Crucially, hedging ESG risks in this manner carries the risk of financial underperformance as long as mitigation measures are deferred (Andersson et al., 2016). If a portfolio of high-ESG stocks from one rating provider outperforms the high-ESG portfolios of other providers, the former would represent a cheaper hedge and be attractive for investors seeking to insure against ESG risks.

The ultimate aim for investors solely concerned with the financial performance of their portfolio is to construct portfolios with an optimal risk-return trade-off. A standard tool for constructing such portfolios is the mean-variance optimization by Markowitz (1952). However, when applied to more than two assets in real-world portfolio allocation tasks, mean-variance optimization usually performs poorly due to estimation errors (Michaud, 1989) and corner solutions (Broadie, 1993). From an investor perspective, there are several approaches to handle these problems. First, if investors are unsure about the expected return and risk of several assets, it is most convenient and sufficiently efficient to either buy the CAPM market portfolio (Markowitz, 1952; Sharpe, 1964) or to buy a naïve 1/N portfolio (DeMiguel et al., 2009). In these approaches, the assets and their returns are either weighted by their market capitalization or equally weighted, respectively. Another approach that does not require risk and return forecasts is minimum variance portfolio optimization. The minimum variance portfolio is optimal for investors who favor low return volatility. Since some investors are willing to give up some return to hedge ESG risk (Bolton & Kacperczyk, 2021), their preferences align with the features of a minimum variance portfolio. According to DeMiguel et al. (2009), minimum variance portfolio optimization is the only optimization model that sometimes outperforms the naïve 1/N portfolio.

In contrast, Pedersen, Babu, and Ari (2021) develop an enhanced method that builds on mean-variance optimization and find that this enhanced approach yields significant alpha beyond the market and the 1/N portfolio. Moreover, Bessler and Wolff (2024) use an out-of-sample Black-Litterman portfolio optimization framework and show that portfolio optimization with sector return prediction models significantly outperforms portfolios using historical averages as well as passive benchmark portfolios. These results indicate that sophisticated portfolio optimization methods can outperform the market. However, these approaches require return expectations as a proxy for investors' opinions.

We are not aware of a dataset that contains return expectations of similar quality for several stock markets around the globe. Therefore, we focus on three investor perspectives. First, an investor who implements a naïve diversification strategy by equally weighting the stocks in the portfolio. Second, an investor who follows a market-oriented strategy by weighting stocks according to their market value. Third, an investor who follows a risk-mitigating approach by using a minimum-variance optimization model.

## 3. ESG ratings

An essential prerequisite for forming ESG portfolios is information about a company's ESG risks. These risks are assessed by ESG rating agencies that provide ESG scores as a guide for investors (Renneboog et al., 2008; Riedl & Smeets, 2017). ESG risks can materialize as systematic risks (e.g., "greenium")<sup>4</sup> or idiosyncratic stock price risk (e.g.

<sup>4</sup> Pástor et al. (2021); Pedersen et al. (2021); Zerbib (2022).

Horn, 2023). ESG rating agencies and their customers assume that companies with high ESG ratings, i.e., the most sustainable companies, may face lower exposure to idiosyncratic risks such as climate-related transition risks or governance-related scandals.

The relationship between ESG ratings and stock performance has been analyzed in a vast number of studies.<sup>5</sup> In summary, these studies find more support for a non-negative relationship between ESG ratings and stock performance than for a negative relationship, indicating that more sustainable stocks yield higher risk-adjusted returns (Friede et al., 2015; Liang & Renneboog, 2020). These findings contrast with the predictions of asset pricing models by Pástor et al. (2021), Pedersen, Fitzgibbons, and Pomorski (2021), and Zerbib (2022). Moreover, studies support the *raison d'être* of ESG ratings: a significant negative relationship between ESG ratings and idiosyncratic risk<sup>6</sup> as well as downside risk.<sup>7</sup> However, the magnitude of the negative relationship between idiosyncratic risk and ESG ratings is statistically significant but not economically meaningful (Horn, 2023; Oehler & Horn, 2021).

An important but often ignored aspect in early studies is that ESG ratings are provided by several providers employing different rating approaches. A brief overview of the approaches used by ISS oekom, MSCI KLD, Sustainalytics, Refinitiv (previously Asset 4), and Vigeo Eiris is provided in Table 1. In our empirical analysis, we exclusively utilize the overall rating scores provided by these agencies. The heterogeneity in ESG rating approaches can be attributed to differences in emphases among them.<sup>8</sup> While ISS oekom has a stronger focus on the supply chain, Sustainalytics targets future ESG issues, and Vigeo Eiris emphasizes corporate social responsibility. Moreover, the weighting of indicators is heterogeneous: Refinitiv uses industry-dependent weights for the E and S pillars. MSCI KLD provides performance indicators without a specific weighting scheme, whereas ISS oekom, Sustainalytics, and Vigeo Eiris employ varying weights for all indicators depending on the industry. In contrast to credit ratings, this heterogeneity in ESG rating approaches often results in rating dispersion, meaning significantly different ESG ratings for the same company.<sup>9</sup>

Few previous studies have analyzed the influence of ESG rating dispersion on stock returns and risk in the cross-section.<sup>10</sup> However, the question of whether portfolio performance differs when investors apply different ESG ratings for sorting and selecting stocks remains severely understudied. We are aware of only one study by Billio et al. (2021) that addresses this question. We extend upon their work by analyzing a global sample of stocks that includes all stocks rated by at least two rating providers, using the remaining stocks as a benchmark to address selection bias.

ESG ratings typically change on an annual basis. Some ESG rating data is stored annually, while others are in monthly datasets. One dataset provides daily updates but only when there is a change in a stock's rating. Some rating providers incorporate negative screens into their overall score, which decreases the score when a company is subject to a negative screen. To standardize the ratings for comparability, we separate the negative screen scores from the overall rating, resulting in what we refer to as the *ESG rating* for better readability. To numerically standardize the ESG ratings, we follow Berg et al. (2022), Oehler and

**Table 1**  
Description of rating approaches.

Rating provider	Brief description of rating approach
ISS oekom	The ESG Corporate Ratings consider ESG risks, opportunities, and impacts along the entire corporate value chain, including a company's supply chain (ISS ESG (2023)). The ratings explicitly dedicate some components of the rating to the sustainable development goals of the United Nations. These components measure the positive and negative impacts of the rated company's products and/or services. Controversies and violations of global norms are also considered
MSCI KLD	The ratings are based on positive and negative performance indicators (MSCI (2016)). The indicators are grouped in seven categories: one environmental, one governance, four social, and one controversial business involvement category. Positive performance indicators are designed to capture management best practices. Negative performance indicators show violations of global conventions, national or international laws, or regulations. All indicators have binary values: "1" if a company meets the assessment criteria; "0" if the company does not
Refinitiv (now called LSEG)	The rating is based on about 630 company-level ESG measures (LSEG (2022)). However, not all measures are applied on all companies. Depending on the industry, a company rating consists of less than 200 measures per industry. The rating provider computes individual E, S, and G scores that are combined in an ESG score. The G score has the same weights across all industries. The weights for the E and S scores vary. In addition, an ESG controversies score is provided
Sustainalytics	Individual E, S, and G scores are computed (Sustainalytics (2019)). The tree pillars are weighted differently across industries. The aim of the rating provider is to indicate future ESG risk. Therefore, the rating includes issues that might be material and typical for the business model and business environment of a company in the future such as carbon emissions and carbon intensity
Vigeo Eiris	The ESG rating measures the integration of ESG factors into companies' managerial systems (Vigeo Eiris (2020)). The rating is based on over 300 principles of action that are derived from conventions, recommendations, and guidelines from international organizations such as the United Nations, the OECD, etc. Per industry sector, only a subset of these principles of action is employed to compute the E, S, and G pillar scores as well as the combined ESG rating

Horn (2021), and Serafeim and Yoon (2022), applying normalized ratings with a mean value of 0 and a standard deviation of 1. This normalization does not alter the rankings when sorting stocks by their ESG rating. However, using normalized ratings enhances comparability with findings from previous studies. The normalized value (z-score) of the ESG rating of stock  $s$  assigned by rating provider  $j$  in month  $m$  is computed with the following formula:

$$ESGRating_{Z,s,j,m} = \frac{ESGRating_{s,j,m} - \overline{ESGRating}_{j,m}}{\sigma_{ESGRating_{j,m}}} \quad (1)$$

With  $ESGRating_{s,j,m}$  as the ESG rating of stock  $s$  assigned by rating provider  $j$  in month  $m$ ,  $\overline{ESGRating}_{j,m}$  as the mean and  $\sigma_{ESGRating_{j,m}}$  as the standard deviation of the ESG ratings assigned by rating provider  $j$  in month  $m$ .

Higher values of  $ESGRating_{Z,s,j,m}$  indicate better ESG-performance, i.e., lower unmanaged ESG risk and, hence, more sustainable companies. As Berg et al. (2022) show that the ESG ratings hardly depend on industry sectors, we do not adjust the ratings for industry sectors.

<sup>5</sup> See Liang and Renneboog (2020) for a review, see also Giese et al. (2019).

<sup>6</sup> Mishra and Modi (2013); Becchetti et al. (2015); Sassen et al. (2016); Bouslah et al. (2013), (2018); Dunn et al. (2018); Giese et al. (2019); Gougler and Utz (2020); Horn (2023); and Monti et al. (2019).

<sup>7</sup> Lins et al. (2017); Albuquerque et al. (2019), (2020); Hoepner et al. (2020); Ding et al. (2021); Pástor and Vorsatz (2020).

<sup>8</sup> For a detailed analysis of the reasons for the divergence between ESG ratings we refer to Berg et al. (2022) and Billio et al. (2021).

<sup>9</sup> Dorfleitner et al. (2015); Chatterji et al. (2016); Kotsantonis and Serafeim (2019); Billio et al. (2021); Berg et al. (2021, 2022); Oehler and Horn (2021).

<sup>10</sup> Gibson et al. (2021); Oehler and Horn (2021); Avramov et al. (2022); Serafeim and Yoon (2022); Berg et al. (2021).

## 4. Data and methodology

### 4.1. Stock data

We use survivorship-bias-free daily total return and stock price data from Thomson Reuters Datastream starting from the beginning of 2014 until the end of 2019. The analysis excludes stock data before 2014 due to the limited availability of ESG ratings prior to that period. The study encompasses all stocks listed in the following indices (with respective geographic regions in parentheses): S&P/ASX Australia All Ordinaries, Hang Seng, S&P/NZX 50, Straits Times (representing Asia-Pacific excluding Japan), MSCI Europe All Cap (Europe), TOPIX (Japan), MSCI North America All Cap (North America including US/Canada).<sup>11</sup> We adopt Ken French's geographical regions<sup>12</sup> to classify stocks into "Europe," "Japan," "Asia Pacific ex Japan," and "North America". This approach offers several advantages. Firstly, using Ken French's factors ensures that any significant findings are not solely attributable to specially calculated factors. Second, it enhances the replicability of our results. Third, due to relatively small sample sizes for Singapore and Hong Kong, grouping these with other Asia-Pacific countries mitigates potential outlier-driven results.

Stock data and ESG rating data are matched by ISIN or ticker symbol if the ISIN was not originally included. To analyze industry sector exposures across different portfolios, we utilize the Global Industry Classification Standard (GICS) by MSCI and S&P, which consists of eleven sectors and is also employed by some rating providers (e.g., ISS oekom). To enhance robustness, sector exposures are provided only for North American portfolios, as portfolios from other geographical regions may contain insufficient stocks per industry sector for reliable analysis.

### 4.2. Methodology

#### 4.2.1. Portfolio construction

Our focus is on stocks that received ESG ratings from at least two different providers in a month  $m$  as we analyze these ratings pairwise. Specifically, we compare ISS oekom ratings with MSCI KLD ratings, ISS oekom ratings with Refinitiv ratings, and so forth, resulting in ten ESG rating pairs due to five rating providers. For each of these pairs, we only include stocks that are rated by both providers, thereby eliminating the influence of differing stock universes covered by the providers within each pairwise comparison. However, this approach results in varying numbers of stock-month observations among the ten rating pairs. To mitigate concerns regarding potential selection bias, we aggregate stocks that receive only one or no ESG rating into an additional portfolio. We then compare the performance of this portfolio with the portfolios constructed from the rating pairs. It's important to note that ESG rating agencies typically focus on the largest companies in a stock market universe, so stocks with only one or no ESG rating are not necessarily

<sup>11</sup> Stocks with tickers C:SMU.UN, C:WRG, US:ACER, US:AMEH, US:APDN, US:ARWR, US:BLNK, US:ELOX, US:HROW, US:IDEX, US:IDRA, US:LTVN, US:LLEX, US:MAMS, US:OCAT, US:OTRK, US:PLX, US:PZZ, US:RIBT, US:SAUC, US:SRNE, US:SVRA, US:TEUM, US:TXMD, US:VTNR, US:XELB, and US:ZYXI are dropped because of data errors. To further minimize the influence of outliers and data errors, we compute the monthly idiosyncratic risk based on daily returns for each stock by running equation (2) (see section 2.3) for each stock for all daily observations for each month and drop the percent of observations with the highest idiosyncratic risk. When we would additionally drop stock-month observations with less than 17 daily observations per month, our dataset decreased by about a half percent of the stock-month observations. However, since our results do not change considerably when we leave these observations in the dataset and because it is not possible to know ex-ante when stocks will not be traded on at least 17 days per month, we decided to leave these observations in the dataset.

<sup>12</sup> [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html#Developed](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed)

less sustainable but often have smaller market capitalizations.

For each rating pair, we construct ten quintile portfolios per geographic region and month  $m$ . The first five portfolios sort stocks based on the ESG ratings from the first provider A, dividing stocks into quintiles within each geographic region every month. Portfolios are formed using quintile values from the previous month, with each quintile representing stocks of differing sustainability levels according to provider A. The next five portfolios are constructed similarly but based on the ESG ratings from the second provider B. Therefore, theoretically, we construct 400 portfolios per month (four geographic regions \* ten rating pairs \* ten portfolios). However, due to insufficient ratings from Sustainalytics for stocks in the Asia-Pacific (ex-Japan) region, our analysis actually covers 360 portfolios from 36 rating pairs (ten rating pairs each in Europe, Japan, and North America; six rating pairs in Asia-Pacific). The methodology for constructing these quintile portfolios is detailed in Fig. 1. In addition to the quintile portfolios, we also compute quartile and sextile portfolios as robustness checks.

#### 4.2.2. Portfolio comparison

To better integrate with the existing literature, we begin with some additional analyses beyond comparing the 360 portfolios pairwise. In the initial part of our analysis, we aggregate the results from these 360 portfolios per geographic region and quintile, evaluating their Sharpe ratio, alpha, and idiosyncratic risk. In the subsequent part of our study, we employ a pairwise approach as depicted in Fig. 1. This involves analyzing the overlap of stocks and differences in the performance between portfolios from each rating pair. For instance, portfolios labeled "1st quintile  $A \cap B$ ; sorted by A" and "1st quintile  $A \cap B$ ; sorted by B" represent portfolios from a single rating pair ( $A \cap B$ ). Similarly, portfolios such as "2nd quintile  $A \cap C$ ; sorted by A" and "2nd quintile  $A \cap C$ ; sorted by C" are also analyzed. The intersection of quintile portfolios across all rating pairs is evaluated over the entire observation period.

While our pairwise approach eliminates discrepancies arising from rating coverage issues, it also results in a smaller universe of stocks for individual portfolio sorting. This limitation is a common challenge for investors adopting an ESG investment strategy. Such strategies often restrict investable stocks, for instance, by focusing solely on the most sustainable companies within a peer group (best-in-class approach) (see Oehler et al. (2018)), and due to incomplete ESG ratings coverage across all stocks. Consequently, ESG-oriented investors may encounter difficulties in achieving diversification and could face significant exposure to idiosyncratic risk,<sup>13</sup> potentially leading to underperformance.<sup>14</sup> Hence, we incorporate a measure of idiosyncratic risk into our analysis.

#### 4.2.3. Performance measurement

The performance of the 360 portfolios is evaluated using the Sharpe ratio (SR) and factor models. To accommodate our three investor perspectives, portfolio returns are computed based on equally weighted returns of the portfolio constituents, market capitalization weighted returns, and weights derived from minimum variance portfolio optimization. To determine the asset weights for a minimum variance portfolio in month  $t + 1$ , we conduct a minimum variance portfolio optimization using the `mport` package in Stata (Dorantes, 2013) in month  $t$ . Short sales are not permitted in this optimization, reflecting our belief that the majority of ESG investors do not engage in short selling. Additionally, the weight assigned to each stock is capped at 20 % to prevent extreme stock weights, known as corner solutions (Bessler et al., 2021). We compute the SR following Sharpe (1994) as

<sup>13</sup> Barnett and Salomon (2006); Geczy et al. (2005); Statman (2000); Lean and Pizzutilo (2020).

<sup>14</sup> Levy (1978); Adler and Kritzman (2008); Ang et al. (2006, 2009); Stambaugh et al. (2015).

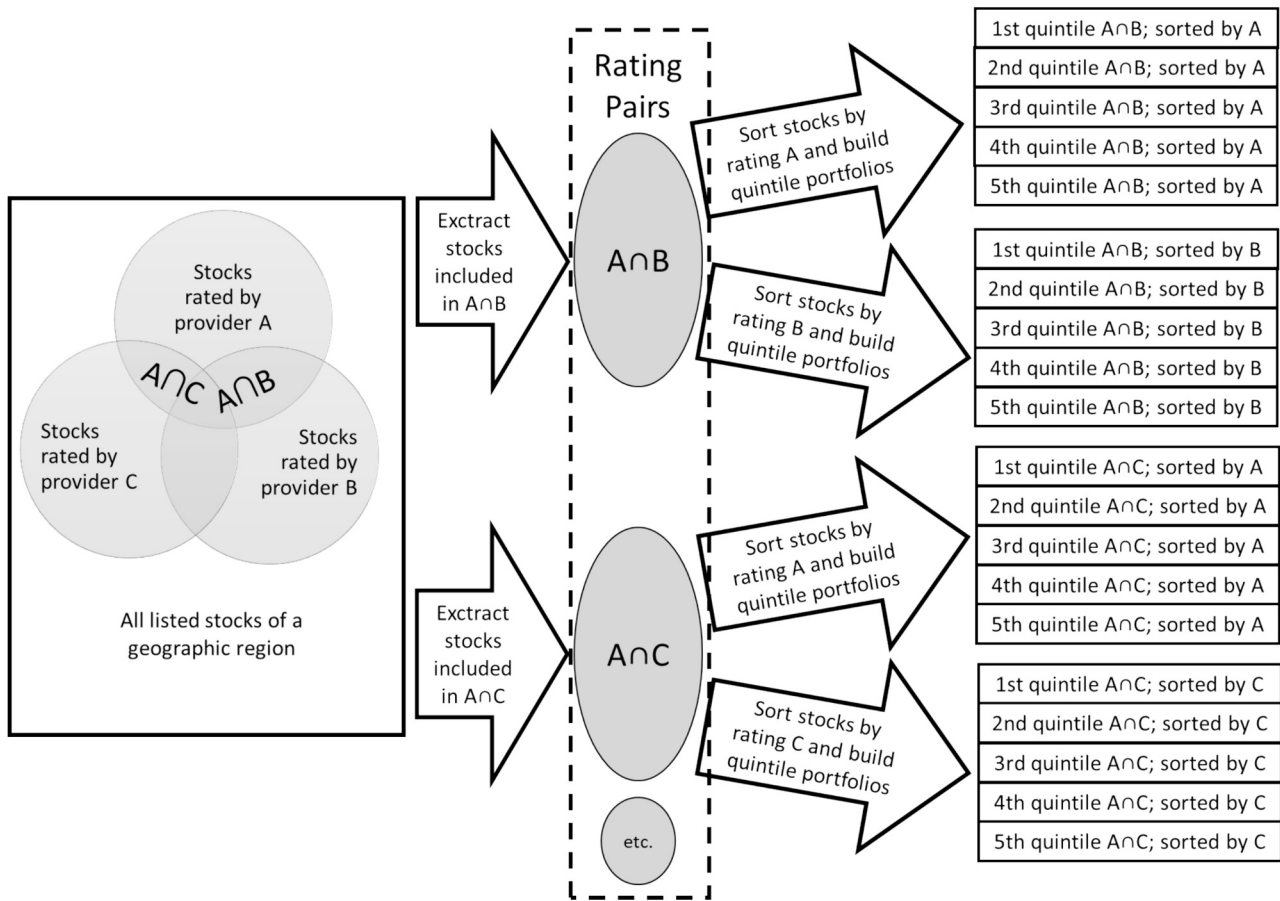


Fig. 1. Methodology to construct quintile portfolios.

$$SR_{it} = \frac{R_{it} - R_{Ft}}{\sigma_{it}} \tag{2}$$

where  $R_{it}$  is the daily mean return of portfolio  $i$  and  $R_{Ft}$  is the daily risk-free return in time period  $t$ .  $\sigma_{it}$  is the standard deviation of the portfolio returns in period  $t$ . Please note that we compute daily SRs per month  $t$ . These SRs usually cannot be annualized by simple multiplication (see Lo (2002)).

To ensure comparability of our findings with the recent literature, we compute alphas and idiosyncratic risk with the Fama and French (2015) five-factor model, which is defined as

$$R_{it} - R_{Ft} = \beta_{1i} * R_{Mt} + \beta_{2i} * SMB_t + \beta_{3i} * HML_t + \beta_{4i} * RMW_t + \beta_{5i} * CMA_t + \alpha_i + \varepsilon_{it} \tag{3}$$

where  $R_{Mt}$ ,  $SMB_t$ ,  $HML_t$ ,  $RMW_t$ , and  $CMA_t$  are the five factors defined by Fama and French (2015).  $\alpha_i$  is the risk-adjusted excess return alpha.  $\varepsilon_{it}$  is the residual per time period  $t$ . We use monthly regression windows. Idiosyncratic risk of portfolio  $i$ , measured as daily idiosyncratic volatility per month  $m$  is defined as  $\sqrt{var \varepsilon_{i,d}}$  for all days  $d$  of month  $m$  (see Bouslah et al. (2018); and Horn (2023) for this approach).<sup>15</sup> All results

<sup>15</sup> We abstain from additionally computing idiosyncratic risk with downside measures (lower partial moments, LPM) since Bouslah et al. (2018) report that results for LPM1 through LPM3 are similar to those for the measure we employ.

are presented in percent. The daily factor data for the Asia-Pacific (ex Japan), European, Japanese, and North American stock markets are from Kenneth French's homepage.<sup>16</sup> Due to its more robust predictive power compared to the five-factor model (Diallo et al., 2023), we employ the Fama and French (1993) three factor model, which only includes the factors  $R_{Mt}$ ,  $SMB_t$ ,  $HML_t$ , as a robustness check. Differences in return, Sharpe ratio, alpha, and idiosyncratic risk between two portfolios are analyzed with  $t$ -tests with Welch's (1947) formula.

### 4.3. Descriptive statistics

We present the number of rating-month observations per provider and region in Panel A of Table 2. Please note that rating-month observations that cannot be matched in at least one rating pair are excluded from the statistics in Table 2. The specific count of stocks considered per rating pair and geographic region is illustrated in Figs. A1–A5 in the appendix. These figures also encompass the total available rating-month observations per rating provider and geographic region, inclusive of unmatched rating-month observations. Refinitiv records the highest number of rating-month observations (236,478) in our sample, while ISS has the fewest (166,148). Panel B of Table 2 displays the number of rating-month observations per rating pair. Due to significant expansions in coverage by all rating providers during the observation period, the number of stocks per rating pair and within quintile portfolios steadily increased, resulting in

<sup>16</sup> [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The factor data used is downloadable under “Developed Markets Factors and Returns”. More specifically, we use the daily data “Fama/French Asia Pacific ex Japan”, “Fama/French European 5 Factors”, “Fama/French Japanese 5 Factors”, and “Fama/French North American 5 Factors”.

**Table 2**

Number of rating-month observations, correlations between ESG ratings, and number of stock-month observations per rating pair in the period from January 2014 to December 2019.

Panel A: Number of rating-month observations per provider					
	Asia-Pacific	Europe	Japan	North America	Total
ISS	21,162	54,328	23,682	66,976	166,148
MSCI	11,620	41,971	17,309	144,312	215,212
Refinitiv	19,267	59,035	28,226	129,950	236,478
Sustainalytics	0	43,870	31,787	127,922	203,579
Vigeo Eiris	23,505	57,257	28,458	60,227	169,447
Panel B: Number of stock-month observations per rating pair					
ISS_MSCI	6593	31,627	16,009	64,568	118,797
ISS_Refinitiv	13,062	47,152	23,371	60,049	143,633
ISS_Sustainalytics	0	33,447	23,595	57,388	114,965
ISS_Vigeo Eiris	20,382	47,810	22,595	48,271	139,058
MSCI_Refinitiv	11,588	37,278	16,870	126,005	191,740
MSCI_Sustainalytics	0	38,327	17,212	125,290	180,828
MSCI_Vigeo Eiris	8116	33,601	16,803	57,554	116,074
Refinitiv_Sustainalytics	0	38,342	28,087	115,746	182,805
Refinitiv_Vigeo Eiris	15,375	51,400	24,906	54,512	146,193
Sustainalytics_Vigeo Eiris	0	35,501	28,272	50,500	114,888
Panel C: Pearson correlations between ratings					
	MSCI_Z	Refinitiv_Z	Sustainalytics_Z	Vigeo Eiris_Z	
ISS_Z	0.47	0.62	0.70	0.73	
MSCI_Z	1.00	0.58	0.59	0.56	
Refinitiv_Z	0.58	1.00	0.73	0.70	
Sustainalytics_Z	0.59	0.73	1.00	0.76	

Notes: Panel A shows the number of rating-month observations per provider per region and in total. Panel B shows the number of stock-month observations per rating pair per region and in total. Panel C provides Pearson correlation coefficients between the z-ratings of the different rating providers. All correlations shown in Panel C show a statistical significance at the 1 % level.

an unbalanced panel analysis. For instance, the ISS and Refinitiv rating pair covered approximately 500 North American stocks in January 2014, indicating simultaneous coverage by both providers at that time. By the end of 2019, this number had risen to about 2300.

The Pearson correlation coefficients of the z-scores are reported in Panel C of Table 2. These coefficients are comparable to those reported by Gibson et al. (2021) and Berg et al. (2022), indicating significant discrepancies in the ESG ratings of individual stocks among the rating providers.

## 5. Results

### 5.1. Individual quintile portfolios

We first present the returns and risk of the 360 quintile portfolios individually before proceeding with the analysis of performance differences between portfolios of each rating pair. Performance differences among the portfolios can be attributed to variations in ESG stock risk (as approximated by ESG ratings), stock weighting (equally weighted, market capitalization weighted, or minimum-variance), and geographical differences. Portfolio returns are presented in Table A1 in the appendix.

#### 5.1.1. Sharpe ratios and alphas

Table 3 presents the mean values of the monthly Sharpe ratio and alpha for the quintile portfolios in each geographic region. Fig. 2 shows the distribution of the Sharpe ratios as boxplots. Fig. 3 presents the distribution of the alphas.

Differences between portfolios attributable to differences in ESG ratings are negligible. A comparison of the Sharpe ratios between fifth and first quintile portfolios reveals no clear overall pattern. However, there are some marginal regional differences. Sharpe ratios of fifth quintile portfolios for Asian-Pacific stocks are slightly lower than those of first quintile portfolios, while for North American stocks, the opposite trend is observed. For European and Japanese stocks, Sharpe ratios

among quintile portfolios reveal only minimal differences. Nevertheless, all differences are not statistically significant, suggesting the results are consistent across regions. This research indicates that investors can mitigate ESG risks by investing in stocks with strong ESG scores while achieving a favorable Sharpe ratio.

Among portfolios of stocks with at least two ESG ratings, the stock weighting also has no clear influence. Portfolios with equally and market capitalization weighted returns have positive Sharpe ratios ranging between 0.03 and 0.15. Some minimum-variance portfolios exhibit negative returns with Sharpe ratios ranging from 0.16 to 0.13. Moreover, minimum variance portfolios within the same quintile reveal a wider distribution of Sharpe ratios relative to equally weighted and market capitalization weighted portfolios. However, the differences between these three portfolio types are not statistically significant.

Geographical differences also have no significant influence on the Sharpe ratios. However, investors focusing solely on maximizing Sharpe ratios may consider investing in market capitalization weighted portfolios consisting of stocks with one or no ESG ratings. Such portfolios often exhibit higher Sharpe ratios compared to most quintile portfolios across all geographic regions. This could be interpreted as a hint of a premium for ESG rating uncertainty (Avramov et al., 2022), although statistically significant effects are observed only in few cases.<sup>17</sup> However, when applied to portfolios of stocks with one or no ESG ratings, the performance of minimum variance and equally weighed portfolios do not indicate exploitable investment opportunities.

The mean monthly alpha across all equally weighted quintile portfolios is 0.27 %. The best performing portfolio yields an alpha of 1.35 % per month, while the worst performing portfolio has an alpha of 0.18 %.

We do not find a clear tendency regarding the relation between ESG rating quintiles and alphas. Overall, the alphas of the equally weighted

<sup>17</sup> Detailed results for statistical significance are available from the authors upon request.

**Table 3**  
Quintile portfolios' Sharpe ratio and alpha.

Quintile	Sharpe ratio			Alpha		
	Mean			Mean		
	Equally weighted	Mcap weighted	Min variance optimization	Equally weighted	Mcap weighted	Min variance optimization
<i>Panel A: Asia-Pacific</i>						
1	0.08	0.08	0.04	0.83	1.15	0.35
2	0.07	0.07	0.01	0.60	0.81	0.58
3	0.06	0.07	0.03	0.49	0.70	0.32
4	0.05	0.06	0.01	0.32	0.60	0.13
5	0.05	0.04	0.05	0.17	0.14	0.60
5-1	0.03	0.04	0.01	0.65	1.01	0.95
≤1 rating	0.09	0.13	0.28	0.84	1.70	1.64
<i>Panel B: Europe</i>						
1	0.06	0.08	0.04	0.12	0.51	0.92
2	0.06	0.08	0.07	0.15	0.50	1.11
3	0.05	0.07	0.03	0.09	0.33	0.01
4	0.06	0.07	0.02	0.12	0.30	0.54
5	0.06	0.07	0.01	0.13	0.41	0.55
5-1	0.00	0.01	0.04	0.01	0.09	1.46
≤1 rating	0.02	0.11	0.01	0.26	0.94	0.69
<i>Panel C: Japan</i>						
1	0.05	0.07	0.05	0.11	0.57	0.60
2	0.05	0.06	0.06	0.21	0.34	1.14
3	0.06	0.07	0.06	0.46	0.68	1.59
4	0.06	0.07	0.04	0.50	0.66	0.67
5	0.06	0.07	0.01	0.51	0.64	0.10
5-1	0.01	0.00	0.04	0.41	0.07	0.70
≤1 rating	0.06	0.10	0.14	0.10	0.83	0.98
<i>Panel D: North America</i>						
1	0.08	0.13	0.05	0.32	0.66	1.16
2	0.08	0.12	0.05	0.13	0.40	0.22
3	0.08	0.13	0.04	0.11	0.30	0.46
4	0.09	0.13	0.07	0.16	0.28	0.98
5	0.10	0.14	0.07	0.23	0.29	0.48
5-1	0.03	0.01	0.02	0.09	0.38	0.68
≤1 rating	0.03	0.14	0.03	0.10	1.08	0.54

*Notes:* The table provides the mean of the monthly Sharpe ratio and alpha of the quintile portfolios of the 36 rating pairs and the portfolio of stocks with only one or no ESG rating (≤1 rating) per geographic region in the period from January 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. The Sharpe ratio of a portfolio is computed with Eq. (2). Portfolio alpha is computed by running Eq. (3) on the daily portfolio returns of every month. We multiply the respective alpha with the number of observations per month to compute the monthly alpha. We compute the Sharpe ratio and alpha for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we take the latter values and compute their mean values. In addition, we compute differences between the fifth quintile portfolios and the first quintile portfolios of all rating pairs. We report the respective mean, minimum, and maximum values of these differences in the lines "5-1". Example: When Sharpe ratios are computed with equally weighted returns, on average, the mean monthly Sharpe ratio of the first quintile portfolios covering North American stocks is 0.08. The difference between the mean Sharpe ratio of the fifth quintile portfolios and the first quintile portfolios is 0.03 on average.

and market capitalization weighted fifth quintile portfolios are lower than those of the first quintile portfolios, although these differences are not statistically significant (see Larcker et al. (2022); Starks (2021)). Comparing alphas across different quintile portfolios constructed with minimum variance optimization reveals higher alphas for portfolios with low ESG ratings. The difference between the alphas of the fifth and first quintile portfolios is negative in most cases, with statistical significance observed in a few instances at the 1 % level. This indicates that the stock weighting algorithm has an influence on the relationship between ESG ratings and alphas.

Minimum variance optimization results in the widest distribution of returns, with alphas that differ significantly from those of equally weighted and market capitalization weighted portfolios but with no clear overall direction. Most market capitalization weighted portfolios show positive alphas, which on average are higher than those of the equally weighted portfolios. However, the alphas of equally weighted and market capitalization weighted portfolios do not differ significantly. This suggests that neither investors employing a naïve diversification

strategy nor those following a market-oriented strategy would have a clear advantage. Retail investors using a naïve diversification strategy can implement it independently, while those employing a market-value-oriented strategy may opt for value weighted index funds with minimal fees. Accounting for these fees, the alphas of equally weighted and market capitalization weighted portfolios are virtually identical.

The relationship between alphas of equally weighted and market capitalization weighted portfolios and ESG rating quintiles also differs among geographical regions. In the samples of Asia-Pacific and North American stocks, the portfolios' alphas are negatively related to ESG ratings (see also Engle (2024)). Conversely, in the sample of Japanese stocks, we observe the opposite trend. Alphas appear insensitive to ESG ratings in the European sample.

These findings support our hypothesis that long-short portfolios based on ESG ratings may outperform the market in some regions while underperforming in others, resulting in average performance that does not exceed market benchmarks. Furthermore, our observation helps explain the mixed findings in the literature regarding the relationship

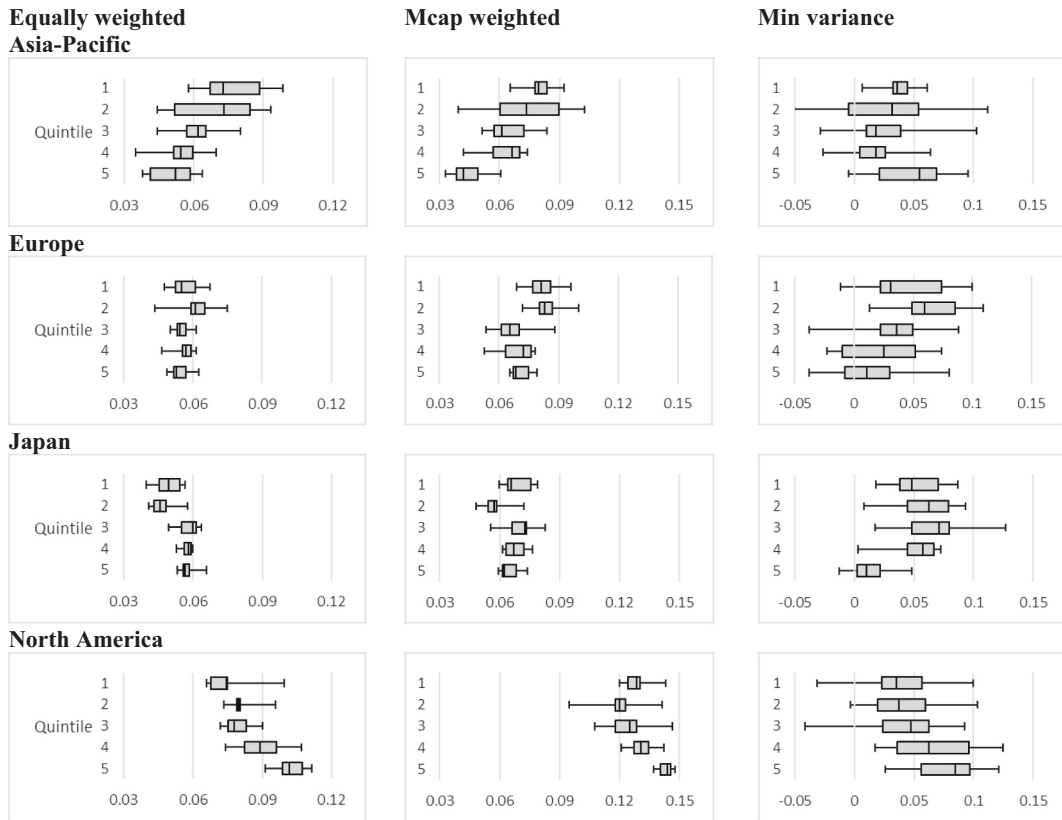


Fig. 2. Boxplot of quintile portfolios' Sharpe ratio.

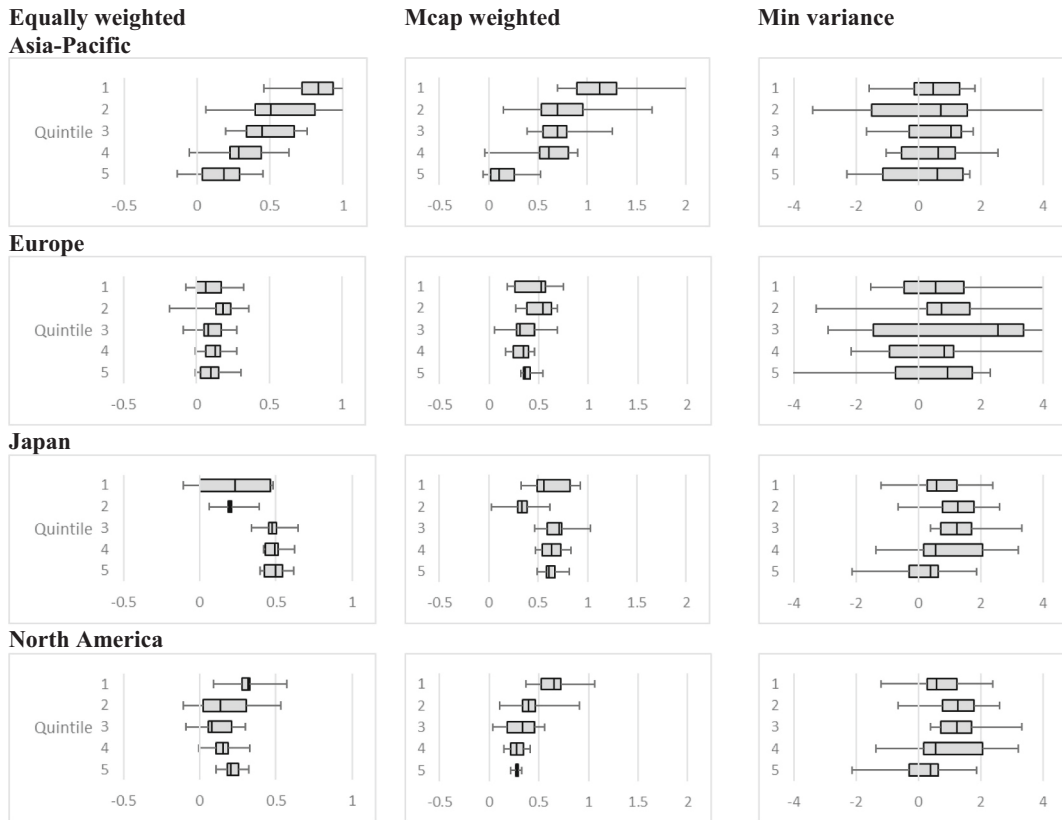


Fig. 3. Boxplot of quintile portfolios' alpha (in percent).

between ESG ratings and stock returns. Additionally, the lack of clear outperformance by high-ESG portfolios, coupled with the systematic risk taken by ETF investors (Ben-David et al., 2018), may explain why investors are withdrawing funds from ESG investments. Moreover, the high alphas of first quintile minimum variance portfolios match the observation by DeMiguel et al. (2009) that minimum variance optimized portfolios outperform equally and market capitalization weighted

**Table 4**  
Quintile portfolios' idiosyncratic risk.

Quintile	Idiosyncratic Risk		
	Mean		
	Equally weighted	Mcap weighted	Min variance optimization
<i>Panel A: Asia-Pacific</i>			
1	0.45	0.50	1.11
2	0.46	0.50	1.00
3	0.42	0.45	0.99
4	0.41	0.42	0.92
5	0.40	0.37	0.85
5-1	0.06	0.13	0.26
≤1 rating	0.32	0.41	0.33
<i>Panel B: Europe</i>			
1	0.32	0.33	0.97
2	0.32	0.32	0.93
3	0.32	0.31	0.90
4	0.30	0.28	0.86
5	0.29	0.27	0.80
5-1	0.03	0.06	0.17
≤1 rating	0.23	0.31	0.60
<i>Panel C: Japan</i>			
1	0.35	0.36	0.97
2	0.32	0.34	0.93
3	0.31	0.33	0.89
4	0.32	0.31	0.91
5	0.31	0.30	0.89
5-1	0.04	0.06	0.08
≤1 rating	0.26	0.29	0.40
<i>Panel D: North America</i>			
1	0.16	0.19	0.94
2	0.14	0.17	0.87
3	0.13	0.15	0.85
4	0.13	0.12	0.82
5	0.13	0.10	0.75
5-1	0.03	0.09	0.20
≤1 rating	0.13	0.15	0.67

*Notes:* The table provides the mean of the daily idiosyncratic risk per month of the quintile portfolios of the 36 rating pairs and the portfolio of stocks with only one or no ESG rating (≤1 rating) per geographic region in the period from January 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). Idiosyncratic risk is computed by first running Eq. (3) on the daily portfolio returns of every month to estimate the betas in Eq. (3). Thereafter, the betas are used to compute the residual  $\varepsilon_{i,t}$  for every observation. Idiosyncratic risk is then computed as the standard deviation of the residuals of every month. We compute the mean daily idiosyncratic risk per month for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we take the latter values and compute their mean values. In addition, we compute differences between the fifth quintile portfolios and the first quintile portfolios of all rating pairs. We report the respective mean values of these differences in the lines "5-1". Example: When portfolio returns are computed with equally weighted returns, on average, the mean idiosyncratic risk of the first quintile portfolios covering North American stocks is 0.16 %. The difference between the mean idiosyncratic risk of the fifth quintile portfolios and the first quintile portfolios is 0.03 % on average.

portfolios in some settings.

Moreover, a market-cap-weighted portfolio of stocks with only one or no ESG rating outperforms the market. The findings regarding equally weighted and minimum variance optimized portfolios of stocks with one or no ESG ratings are mixed and show no clear pattern.

### 5.1.2. Idiosyncratic risk

Results for portfolio idiosyncratic risk are summarized in Table 4 and Fig. 4. We find further evidence that portfolios composed of stocks with higher ESG ratings exhibit lower idiosyncratic risk (see also Oehler and Horn (2021)). The difference between the fifth and first quintile portfolios is statistically significant at the 1 % level in many cases, supporting the notion that high-ESG stocks are less affected by unexpected news shocks. As demonstrated earlier, portfolios of high ESG stocks do not necessarily suffer from significantly lower alphas. This implies that high-ESG rated portfolios offer the opportunity to earn a similar premium for exposure to systematic risk factors while being less exposed to uncompensated idiosyncratic risk compared to portfolios of lower ESG rated stocks.

Stock weighting has a remarkable influence on idiosyncratic risk. Idiosyncratic risk remains similar whether portfolio returns are computed using equally weighted or market capitalization weighted returns. Minimum variance optimization, however, results in significantly higher idiosyncratic risk. On the one hand, this is quite surprising since minimum variance optimization is designed for minimizing portfolio risk. On the other hand, this effect could be explained by the smaller number of stocks and, hence less diversification, in minimum variance portfolios. The patterns are consistent across all four geographical regions.

The idiosyncratic risk of portfolios covering stocks with one or no ESG rating is not higher than that of quintile portfolios. This indicates that a portfolio of unrated stocks does not necessarily suffer from higher idiosyncratic risk compared to a portfolio of rated stocks. However, Russia's invasion of Ukraine may lead to a temporary increase in global fossil energy consumption until the mid-2020s (International Energy Agency, 2022). This could temporarily reduce transition risks, i.e., expected idiosyncratic risk, for low-ESG stocks and make ESG funds less attractive to investors.

Despite such shocks, we find consistent evidence that portfolios of high ESG stocks benefit from lower idiosyncratic risk.

### 5.1.3. Interim conclusion for individual quintile portfolios

The previous results were observed when analyzing the 360 quintile portfolios independently, and they do not yet reveal differences between portfolios of rating pairs. Moreover, results for minimum-variance optimization appear noisy due to their wide distribution. Nevertheless, we already find support for our first hypothesis that long-short portfolios based on ESG ratings outperform the market in some regions while underperforming in others, resulting in average performance that does not surpass market benchmarks. However, it is important to note that the previously reported range between the minimum and maximum Sharpe ratio, alpha, and idiosyncratic risk of the quintile portfolios may arise from two different sources. First, due to coverage differences, some stocks may be included in a portfolio based on the ratings of one provider but not in a portfolio based on the ratings of another provider. To mitigate this issue, we only analyze differences between two portfolios of the same quintile from one rating pair. That is, we adopt a pairwise approach in the following analysis. This approach allows us to focus on the second source of differences between quintile portfolios: variations in opinion among ESG rating providers. The pairwise approach enables us to test our second hypothesis that no single ESG rating is superior for constructing portfolios; portfolios based on the ESG ratings of one provider do not consistently outperform portfolios based on the ESG ratings of other providers.

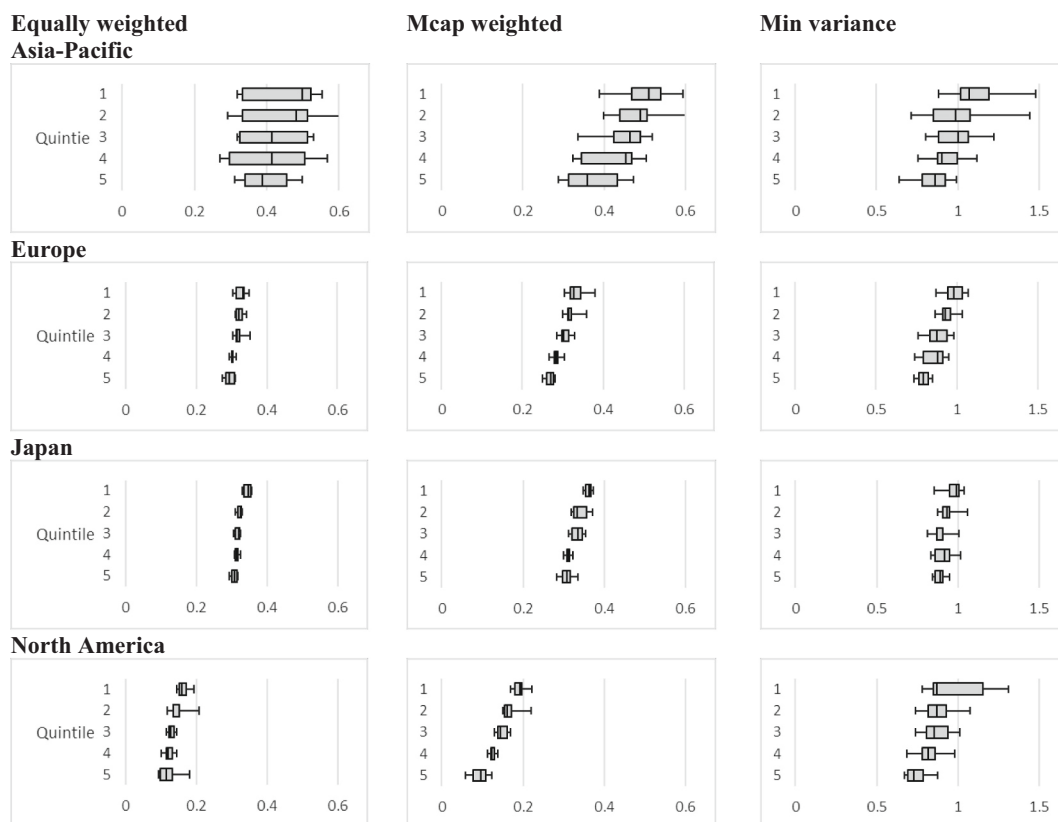


Fig. 4. Boxplot of quintile portfolios' idiosyncratic risk (in percent).

Table 5  
Intersection of quintile portfolios of 36 rating pairs.

Quintile provider A	Quintile provider B				
	1	2	3	4	5
1	52.0 [29.7; 72.3]	25.8 [16.8; 31.1]	13.0 [6.3; 20.3]	6.3 [0.5; 16.2]	2.9 [0.1; 8.4]
2	27.3 [20.3; 37.3]	30.2 [17.4; 39.4]	23.2 [16.4; 29.1]	13.8 [7.3; 24.4]	5.6 [1.4; 12.2]
3	14.5 [6.6; 24.0]	24.4 [20.8; 29.4]	27.5 [16.5; 34.5]	22.1 [16.2; 27.0]	11.4 [5.6; 19.9]
4	6.0 [1.5; 16.6]	14.3 [8.2; 20.6]	23.1 [18.3; 28.0]	29.6 [22.2; 36.7]	26.9 [20.5; 32.0]
5	1.8 [0.0; 6.0]	5.5 [1.2; 10.9]	12.6 [4.9; 18.6]	28.4 [18.0; 33.4]	51.8 [37.9; 66.7]

Notes: The table provides the mean, minimum, and maximum pairwise intersection of the 360 portfolios of the 36 rating pairs in the period from January 2014 to December 2019. The minimum and maximum values of the size of the intersection are shown in the squared brackets. All values are presented in percent. Example: On average, 30.2 % of the stocks included in the second quintile portfolio of rating provider A are also in the second quintile portfolio of rating provider B. The minimum (maximum) values of the size of the intersection for the second quintile portfolios is 17.4 % (39.4 %).

5.2. Pairwise approach

5.2.1. Intersection and industry sector exposure

Table 5 provides an overview of the intersection size of the 360 portfolios across the 36 rating pairs. Specifically, we present the mean, minimum, and maximum percentages of stocks included in corresponding quintile portfolios of each rating pair. If ESG ratings from

different providers were entirely unrelated, the intersection of any two portfolios from different rating providers would comprise 20 % of the stocks in a quintile portfolio. A correlation of 1 between ESG ratings would result in a 100 % intersection for portfolios of the same quintile and 0 % for portfolios of different quintiles. In comparison, we find a relatively small intersection averaging 38.2 % between two portfolios of the same quintile. The intersection varies significantly across different quintiles. For instance, an average of 27.5 % of stocks in the third quintile portfolio of rating provider A also appear in the third quintile portfolio of rating provider B. However, the mean intersection among first quintile portfolios and fifth quintile portfolios is 52 %. The highest intersection among rating pairs is 72.3 % for two portfolios of the first quintile, which represents portfolios with the lowest ESG ratings. These patterns are consistent across portfolios covering Asia-Pacific, European, Japanese, and North American stocks.<sup>18</sup>

Due to the limited overlap in portfolio constituents, it appears plausible that the quintile portfolios are exposed differently to industry sectors. We present plots of the mean industry exposure of North American quintile portfolios per rating provider in Fig. A6 in the appendix. These plots suggest that industry dependencies have minimal influence on the pairwise comparisons, as we either observe similar trends across different ESG ratings or random deviations.

5.2.2. Sharpe ratios and alphas

We analyze differences in Sharpe ratios (see Table 6) and alphas (see Table 7) between portfolios within the same ESG rating quintile across rating pairs. The results strongly support our hypothesis that portfolios based on the ESG ratings of one provider do not outperform portfolios based on the ESG ratings of other providers in any region. The analysis reveals no significant differences in Sharpe ratios and alphas between

<sup>18</sup> See the appendix, Table A2A3 for the intersection per region.

**Table 6**  
Differences between quintile portfolios' Sharpe ratio.

Quintile	Sharpe ratio					
	Equally weighted		Mcap weighted		Min variance optimization	
	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value
<i>Panel A: Asia</i>						
1	0.01 [0.00; 0.01]	77.2	0.00 [0.00; 0.02]	62.5	0.02 [0.00; 0.04]	55.9
2	0.01 [0.00; 0.01]	76.4	0.01 [0.00; 0.03]	52.2	0.07 [0.01; 0.21]	5.4
3	0.01 [0.00; 0.02]	65.2	0.01 [0.00; 0.01]	75.5	0.05 [0.01; 0.09]	9.9
4	0.01 [0.00; 0.02]	60.8	0.01 [0.00; 0.01]	73.3	0.03 [0.00; 0.06]	34.1
5	0.01 [0.00; 0.01]	78.2	0.01 [0.00; 0.01]	75.7	0.02 [0.00; 0.07]	26.4
<i>Panel B: Europe</i>						
1	0.01 [0.00; 0.02]	70.3	0.01 [0.00; 0.03]	48.6	0.04 [0.00; 0.08]	15.8
2	0.01 [0.00; 0.02]	57.4	0.01 [0.00; 0.02]	62.9	0.04 [0.01; 0.09]	10.7
3	0.01 [0.00; 0.01]	89.2	0.01 [0.00; 0.03]	39.1	0.03 [0.01; 0.12]	1.9
4	0.01 [0.00; 0.01]	75.1	0.01 [0.00; 0.02]	55.5	0.03 [0.00; 0.08]	11.2
5	0.01 [0.00; 0.01]	80.3	0.01 [0.00; 0.01]	73.6	0.04 [0.00; 0.10]	6.4
<i>Panel C: Japan</i>						
1	0.00 [0.00; 0.01]	75.5	0.01 [0.00; 0.02]	46.7	0.02 [0.00; 0.05]	30.1
2	0.00 [0.00; 0.02]	53.8	0.01 [0.00; 0.02]	48.1	0.03 [0.01; 0.05]	18.3
3	0.00 [0.00; 0.01]	67.8	0.01 [0.00; 0.02]	50.3	0.02 [0.00; 0.10]	0.9
4	0.00 [0.00; 0.01]	86.0	0.01 [0.00; 0.01]	59.1	0.02 [0.00; 0.06]	12.9
5	0.01 [0.00; 0.01]	66.2	0.01 [0.00; 0.01]	63.0	0.03 [0.00; 0.06]	13.9
<i>Panel D: North America</i>						
1	0.01 [0.00; 0.02]	52.1	0.01 [0.00; 0.02]	56.8	0.04 [0.01; 0.10]	3.4
2	0.01 [0.00; 0.02]	64.2	0.01 [0.00; 0.03]	31.1	0.04 [0.01; 0.09]	9.1
3	0.01 [0.00; 0.02]	66.7	0.01 [0.00; 0.03]	44.2	0.02 [0.00; 0.05]	35.8
4	0.01 [0.00; 0.02]	65.4	0.01 [0.00; 0.02]	62.2	0.03 [0.01; 0.07]	9.7
5	0.01 [0.00; 0.02]	68.0	0.01 [0.00; 0.01]	78.4	0.04 [0.03; 0.07]	16.1

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly Sharpe ratios per quintile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. The Sharpe ratio of a portfolio is computed with Eq. (2). We compute the mean monthly return and Sharpe ratio for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between the mean Sharpe ratios of the two matching quintile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum *p*-value of the *t*-tests per two matching quintile portfolios among the rating pairs. Example: When Sharpe ratios are computed with equally weighted returns, on average, the difference in mean monthly Sharpe ratios between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 0.01 for North American stocks, with a minimum value of 0.00 and a maximum value of 0.02. The smallest *p*-value of the *t*-tests on the difference in return between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 52.1 % for North American stocks. This shows that the monthly returns of the two most different portfolios are not different from each other at statistically significant levels.

portfolios within the same ESG rating quintile. This indicates that when rating providers cover an identical universe of investable stocks, portfolio sorting based on the ESG ratings of any provider leads to quintile portfolios that do not significantly differ in Sharpe ratio and alpha compared to quintile portfolios sorted according to the ESG ratings of other providers. Hence, there is no ESG rating that is more suitable for investors seeking persistent outperformance. Using two or more ESG ratings to sort and select stocks results in portfolios with considerably different constituents depending on the ESG provider. However, the corresponding effect on portfolio Sharpe ratios and alphas is negligible.

**5.2.3. Idiosyncratic risk**

Somewhat surprisingly, we find significant differences in idiosyncratic risk among the quintile portfolios of some rating pairs. In Table 8, we provide the mean, minimum, and maximum of the absolute values of the mean differences in idiosyncratic risk between portfolios within the same quintile among the 36 rating pairs. Furthermore, we report the smallest *p*-value from all *t*-tests conducted among the 36 rating pairs per quintile. We specifically focus on the first and fifth quintile portfolios. Statistically significant differences in idiosyncratic risk are observed among Asia-Pacific and North American stocks, but not in the European and Japanese samples. For the Asia-Pacific stock sample, when returns are computed using equally weighted or market capitalization weighted methods, there is at least one rating pair where the two first quintile portfolios exhibit different idiosyncratic risk levels with statistical significance at the 1 % level. Additionally, significant differences in idiosyncratic risk are found between the two fifth quintile portfolios of certain rating pairs. In the case of North American stocks, differences in idiosyncratic risk between two first quintile portfolios of some rating pairs are significant at the one per mill level. Similar statistically significant differences are also found between portfolios of the fifth quintile. These significant differences can be an indication that the ESG ratings of some providers are stronger related to idiosyncratic risk than the ESG ratings of other providers. A stronger relation would indicate a higher suitability of an ESG rating for predicting ESG stock risk. We further investigate this issue in the following.

**5.2.4. Is one ESG rating superior for predicting idiosyncratic risk?**

Boxplots are used to visualize whether portfolios based on a specific ESG rating provider exhibit significantly lower idiosyncratic risk compared to portfolios based on ratings from other providers. Fig. 5 presents one boxplot per rating provider and ESG rating quintile. These

**Table 7**  
Differences between quintile portfolios' alpha (FF-5-factor model).

Quintile	Alpha (FF-5)					
	Equally weighted		Mcap weighted		Min variance optimization	
	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value
<i>Panel A: Asia</i>						
1	0.11 [0.00; 0.42]	38.1	0.23 [0.03; 0.76]	14.8	0.87 [0.43; 1.58]	53.3
2	0.22 [0.06; 0.34]	29.9	0.34 [0.03; 0.76]	8.7	2.23 [0.15; 4.31]	4.3
3	0.20 [0.01; 0.52]	28.4	0.27 [0.04; 0.46]	30.9	1.22 [0.40; 2.52]	15.7
4	0.28 [0.02; 0.68]	13.2	0.24 [0.13; 0.36]	39.7	1.30 [0.53; 2.77]	20.5
5	0.24 [0.14; 0.40]	31.6	0.11 [0.04; 0.18]	50.6	0.95 [0.23; 2.01]	22.0
<i>Panel B: Europe</i>						
1	0.10 [0.02; 0.27]	34.5	0.20 [0.01; 0.49]	7.4	2.76 [0.52; 11.63]	1.6
2	0.17 [0.01; 0.49]	7.5	0.21 [0.05; 0.38]	17.5	2.09 [0.05; 5.22]	1.9
3	0.08 [0.00; 0.20]	40.0	0.17 [0.03; 0.45]	8.1	1.88 [0.12; 5.28]	0.3
4	0.07 [0.02; 0.17]	43.8	0.07 [0.03; 0.14]	54.2	1.78 [0.01; 11.14]	23.2
5	0.07 [0.04; 0.16]	47.3	0.09 [0.02; 0.15]	38.5	1.60 [0.18; 4.17]	15.9
<i>Panel C: Japan</i>						
1	0.13 [0.03; 0.25]	34.1	0.16 [0.01; 0.40]	17.4	1.35 [0.22; 3.59]	10.1
2	0.08 [0.01; 0.18]	40.6	0.14 [0.02; 0.36]	18.3	1.24 [0.35; 2.09]	17.7
3	0.12 [0.00; 0.23]	29.7	0.20 [0.03; 0.32]	23.4	1.17 [0.30; 2.21]	15.6
4	0.07 [0.01; 0.13]	56.6	0.10 [0.01; 0.26]	32.9	1.71 [0.10; 4.57]	5.1
5	0.08 [0.01; 0.20]	33.3	0.10 [0.01; 0.21]	29.0	1.17 [0.32; 2.63]	7.2
<i>Panel D: North America</i>						
1	0.09 [0.01; 0.21]	20.1	0.22 [0.02; 0.54]	2.0	2.45 [0.95; 4.54]	0.8
2	0.17 [0.07; 0.33]	3.7	0.21 [0.02; 0.54]	1.2	1.51 [0.02; 3.99]	3.2
3	0.07 [0.00; 0.17]	25.4	0.14 [0.01; 0.31]	5.2	1.50 [0.37; 4.06]	15.1
4	0.12 [0.01; 0.26]	8.4	0.09 [0.00; 0.26]	3.7	0.91 [0.10; 1.67]	30.9
5	0.07 [0.00; 0.17]	13.6	0.03 [0.00; 0.07]	55.9	1.43 [0.08; 3.88]	2.7

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly alpha per quintile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). Portfolio alpha is computed by running Eq. (3) on the daily portfolio returns of every month. We multiply the respective alpha with the number of observations per month to compute the monthly alpha. We compute the mean monthly alpha for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between the mean alphas of the two matching quintile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum p-value of the t-tests per two matching quintile portfolios among the rating pairs. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean monthly alpha between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 0.09 % for North American stocks, with a minimum value of 0.01 % and a maximum value of 0.21 %. The smallest p-value of the t-tests on the difference in alpha between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 20.1 % for North American stocks. This shows that the monthly alphas of the two most different portfolios are not different from each other at statistically significant levels.

boxplots illustrate the differences in idiosyncratic risk between portfolios containing stocks from the first (fifth) ESG rating quintile of a specific provider versus those of the remaining providers. Portfolio returns are computed using stock returns weighted by market capitalization.<sup>19</sup>

The boxplots for the first quintile portfolios span from approximately 0.02 % to 0.02 % across all five ESG rating providers. This indicates that first quintile portfolios based on the ratings of any specific provider do not consistently exhibit higher or lower idiosyncratic risk compared to those based on ratings from other providers.

In contrast, the boxplot for fifth quintile portfolios based on Refinitiv ratings consistently shows negative values. This suggests that fifth quintile portfolios based on Refinitiv ratings consistently display lower idiosyncratic risk than those based on ratings from any other provider. However, the absolute magnitude of these differences does not appear economically significant. For instance, the market capitalization-weighted fifth quintile portfolios in North America (Asia-Pacific) show daily idiosyncratic risk ranging from 0.06 % to 0.12 % (0.29 % to 0.47 %). On average, portfolios based on Refinitiv ratings exhibit only a 0.017 % (0.029 %) lower daily idiosyncratic risk compared to portfolios based on other ratings. Furthermore, Refinitiv ratings are known to have undergone rewriting issues (see Berg et al., 2020). Therefore, it cannot be conclusively stated whether portfolios based on the original Refinitiv ratings, before the rewriting, would have also demonstrated lower idiosyncratic risk.

In summary, these findings on idiosyncratic risk provide additional support for our hypothesis that portfolios based on the ESG ratings of one provider do not outperform portfolios based on the ESG ratings of other providers in any region.

A likely reason for this observation is the noise in the measurement of ESG scores, which ultimately reduces differences across the portfolios (see Berg et al., 2021). When ESG ratings are sufficiently noisy, portfolios sorted based on these ratings resemble random stock selections. If these resulting quintile portfolios include enough stocks to achieve a reasonable level of diversification, approaching that of the market portfolio (see Evans & Archer, 1968; Statman, 1987, 2004), the portfolios would exhibit similar return and risk characteristics despite significant discrepancies among different rating providers. Moreover, our

<sup>19</sup> Boxplots for equally weighted portfolios are presented in Fig. A7 in the appendix.

**Table 8**  
Differences between quintile portfolios' idiosyncratic risk (FF-5-factor model).

Quintile	Idiosyncratic risk					
	Equally weighted		Mcap weighted		Min variance optimization	
	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value	Mean [min; max] of mean difference	T-test min p- value
<i>Panel A: Asia-Pacific</i>						
1	0.03 [0.00; 0.07]	0.2	0.06 [0.02; 0.12]	0.0	0.07 [0.01; 0.17]	12.1
2	0.08 [0.02; 0.27]	0.0	0.07 [0.00; 0.25]	0.0	0.18 [0.00; 0.73]	0.0
3	0.01 [0.00; 0.04]	11.7	0.04 [0.01; 0.09]	0.0	0.09 [0.04; 0.19]	0.9
4	0.05 [0.02; 0.08]	0.0	0.04 [0.00; 0.11]	0.0	0.09 [0.02; 0.13]	8.2
5	0.04 [0.04; 0.06]	0.5	0.03 [0.00; 0.05]	1.3	0.09 [0.03; 0.15]	3.2
<i>Panel B: Europe</i>						
1	0.01 [0.00; 0.03]	7.8	0.02 [0.00; 0.04]	2.3	0.06 [0.01; 0.11]	12.0
2	0.01 [0.00; 0.03]	11.7	0.02 [0.00; 0.04]	2.2	0.06 [0.00; 0.11]	22.9
3	0.02 [0.00; 0.04]	1.4	0.01 [0.00; 0.02]	19.6	0.05 [0.00; 0.12]	5.8
4	0.01 [0.00; 0.01]	56.1	0.01 [0.00; 0.03]	7.1	0.03 [0.01; 0.06]	44.7
5	0.01 [0.00; 0.03]	7.6	0.01 [0.00; 0.01]	34.4	0.04 [0.00; 0.08]	12.1
<i>Panel C: Japan</i>						
1	0.01 [0.00; 0.02]	14.7	0.01 [0.00; 0.02]	21.5	0.05 [0.01; 0.09]	11.3
2	0.01 [0.00; 0.02]	31.0	0.02 [0.01; 0.04]	2.0	0.06 [0.00; 0.11]	9.8
3	0.00 [0.00; 0.01]	60.1	0.01 [0.00; 0.03]	8.5	0.06 [0.00; 0.17]	0.9
4	0.01 [0.00; 0.02]	31.9	0.01 [0.00; 0.01]	49.1	0.06 [0.01; 0.12]	7.3
5	0.01 [0.00; 0.02]	29.1	0.01 [0.00; 0.02]	31.5	0.04 [0.01; 0.08]	19.2
<i>Panel D: North America</i>						
1	0.01 [0.00; 0.03]	0.1	0.02 [0.00; 0.03]	0.1	0.06 [0.01; 0.17]	16.6
2	0.02 [0.00; 0.09]	0.0	0.01 [0.00; 0.06]	0.1	0.10 [0.02; 0.23]	3.3
3	0.01 [0.00; 0.02]	0.5	0.01 [0.00; 0.04]	0.0	0.04 [0.01; 0.10]	16.9
4	0.01 [0.00; 0.02]	0.2	0.01 [0.00; 0.02]	0.0	0.07 [0.00; 0.17]	3.3
5	0.02 [0.00; 0.04]	0.0	0.01 [0.00; 0.02]	0.0	0.04 [0.00; 0.17]	4.3

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in daily idiosyncratic risk per quintile portfolio and geographic region of the rating pairs in the period from January 2014 to December 2019. All values are presented in percent. Idiosyncratic risk per portfolio is computed by first running Eq. (3) on the daily portfolio returns of every month to estimate the betas in Eq. (3). Thereafter, the betas are used to compute the residual  $\epsilon_{i,t}$  for every observation. Idiosyncratic risk is then computed as the standard deviation of the residuals of every month. We compute the mean daily idiosyncratic risk per month for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between mean idiosyncratic risk of the two matching quintile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum p-value of the t-tests per two matching quintile portfolios among the rating pairs. Portfolio returns are computed with equally weighted stock returns as well as with stock returns weighted by stocks' market capitalization. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean daily idiosyncratic risk between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 0.01 % for North American stocks, with a minimum value of 0.00 % and a maximum value of 0.03 %. The smallest p-value of the t-tests on the difference in idiosyncratic risk between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 0.1 % for North American stocks. This shows that the idiosyncratic risk of the two most different portfolios is different from each other with a statistical significance at the 1% level.

results support the hypothesis that market prices already reflect the information covered in ESG ratings.

### 5.3. Robustness check

The findings on alphas and idiosyncratic risk remain robust when computed using the Fama and French (1993) three-factor model instead of the five-factor model (see Table A2 in the appendix). As a further robustness check, we form quartile and sextile portfolios instead of quintile portfolios. These results are detailed in Tables A3–A9 in the appendix, revealing a similar pattern to that observed for the quintile portfolios. Despite the small intersection, returns, alphas, and idiosyncratic risk of portfolios within the same quartile/sextile do not differ at economically significant levels. Consequently, we conclude that when applied to the same universe of stocks, portfolios based on the ESG ratings of different providers do not exhibit significant differences in corresponding portfolio Sharpe ratios, alphas, and idiosyncratic risks. Therefore, long-short portfolios do not, on average, outperform the market, and portfolios based on the ESG ratings of one provider do not outperform portfolios based on the ESG ratings of other providers in any region.

## 6. Conclusion

This paper contributes to the literature by comparing the constituents and performance of stock portfolios constructed based on sorts of different ESG ratings. The empirical findings address three important issues for investors, financial analysts, and researchers. First, we confirm the results of previous studies that ratings of different ESG providers differ in their assessments of companies. They also cover different companies due to their mainly focusing on one region. Consequently, quintile portfolios sorted on different ESG ratings contain different constituents, even when the portfolio sorts are based on the same sample of stocks. Second, while long-short portfolios based on ESG ratings may outperform in some markets, they underperform in others, resulting in, on average, no outperformance compared to the broader market. Third, portfolios formed by sorting according to the ESG ratings of one provider lead to quintile portfolios that do not exhibit significant differences in Sharpe ratio, alpha, and idiosyncratic risk compared to the respective

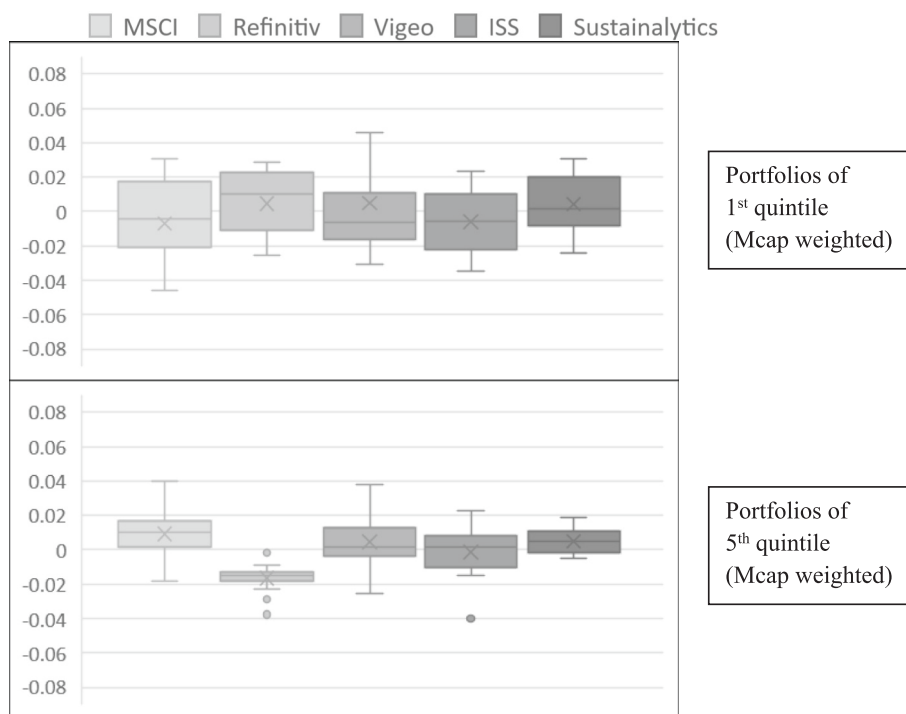


Fig. 5. Differences in daily idiosyncratic risk between portfolios per rating provider in the period from January 2014 to December 2019.

*Notes:* The figure shows one boxplot per rating provider and quintile portfolio. Each boxplot summarizes the values of the differences in daily idiosyncratic risk between the quintile portfolios of the corresponding rating provider and the respective quintile portfolios of the remaining rating providers. All values are presented in percent. We only present the boxplots for the first and fifth ESG rating quintile portfolios of a rating provider. The boxplots include data of the quintile portfolios of the geographic regions Asia-Pacific, Europe, Japan, and North America in the period from January 2014 to December 2019. Idiosyncratic risk per portfolio is computed by first running Eq. (3) on the daily portfolio returns of every month to estimate the betas in Eq. (3). Thereafter, the betas are used to compute the residual  $\varepsilon_{i,d}$  for every observation. Idiosyncratic risk is then computed as the standard deviation of the residuals of every month. We compute the mean daily idiosyncratic risk per month for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between mean idiosyncratic risk of the two matching quintile portfolios of a rating pair. The boxplots show the mean of these differences per rating provider. The figure shows results for portfolios with stock returns weighted by stocks' market capitalization. Results are presented in percent. Example: On average, the difference in mean daily idiosyncratic risk between the fifth quintile portfolios constructed with sorts on ESG ratings of MSCI and the fifth quintile portfolios constructed with sorts on ESG ratings of the remaining rating providers is 0.01 %. The respective differences range between -0.02 % and 0.04 %. This means that the fifth quintile portfolios constructed with sorts on ESG ratings of MSCI show a daily idiosyncratic risk that is between 0.02 % lower and 0.04 % higher than the daily idiosyncratic risk of the fifth quintile portfolios constructed with sorts on ESG ratings of the remaining rating providers.

quintile portfolios based on the ESG ratings of another providers. Therefore, no ESG rating provider offers ESG rating information that investors can use to construct persistently outperforming portfolios.

These findings have important implications for investors and researchers. For investors using only one ESG rating provider, the universe of rated stocks (and, depending on the investment policy, the applied negative screens) is much more important than the approach determining the overall ESG rating. The absence of significant differences between quintile portfolios of different rating providers suggests that significant differences found between portfolios based on the ESG ratings of one provider may also arise between portfolios based on the ESG ratings of another provider when applied to the same universe of stocks. However, this requires testing using a pairwise approach due to the low agreement among rating providers.

There are many other issues that could be addressed in future research. For example, investors with access to multiple ESG ratings could attempt to enhance portfolio performance by incorporating measures of ESG rating disagreement, as differences in opinion may influence the risk and return of individual stocks and stock portfolios (see Avramov et al. (2022); Berg et al. (2021); Gibson et al. (2021);

Oehler and Horn (2021); Serafeim and Yoon (2022)). However, it needs to be noticed that when ratings are compared pairwise, the differences in assessments among ESG rating providers are most likely too low to have a substantial impact at a portfolio level. Further research could investigate whether our findings, based on overall ESG ratings also hold when portfolios are separately formed on the E, S, or G scores. Results may vary, as Bolton and Kacperczyk (2021) demonstrate that certain individual components of ESG scores significantly relate to future stock returns. Additionally, Galema and Gerritsen (2022) highlight the significance of ESG rating revisions, particularly in the environmental dimension. These revisions could reveal important information that could be exploited. As the disclosure of ESG information will increase and be more standardized in the future, this offer the opportunity for more research to investigate the financial benefits of investing in more sustainable stocks. All this is left for future research.

#### Data availability

The authors do not have permission to share data.

Appendix A. Appendix

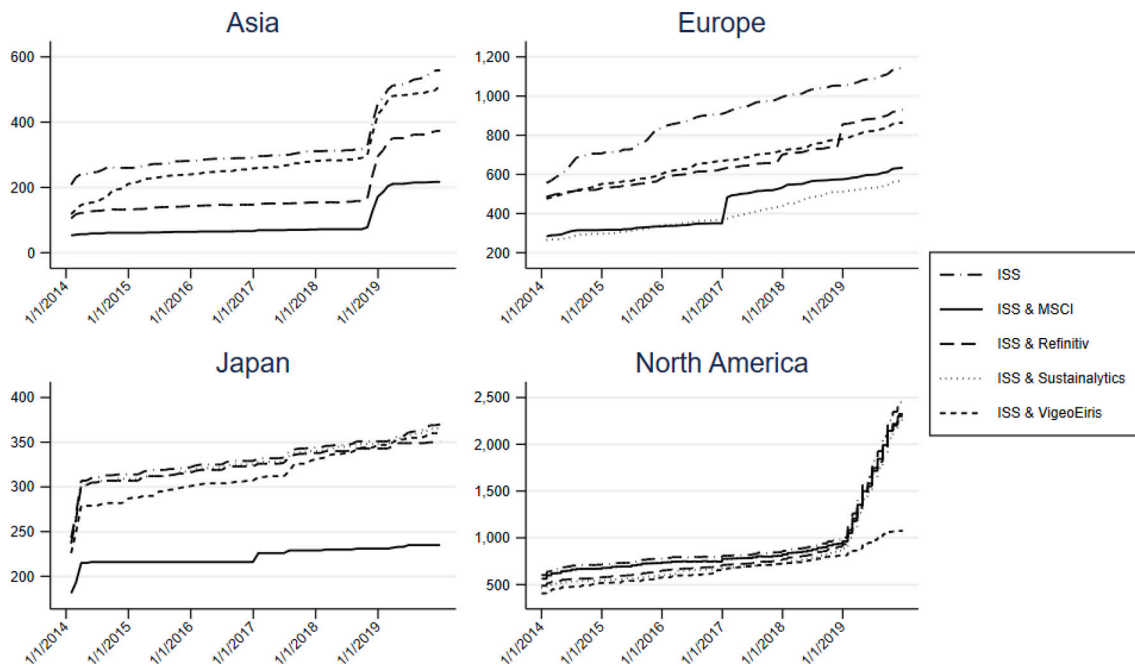


Fig. A1. Number of stocks that are rated by ISS in total and number of stocks considered in rating pairs in the period from January 2014 to December 2019.

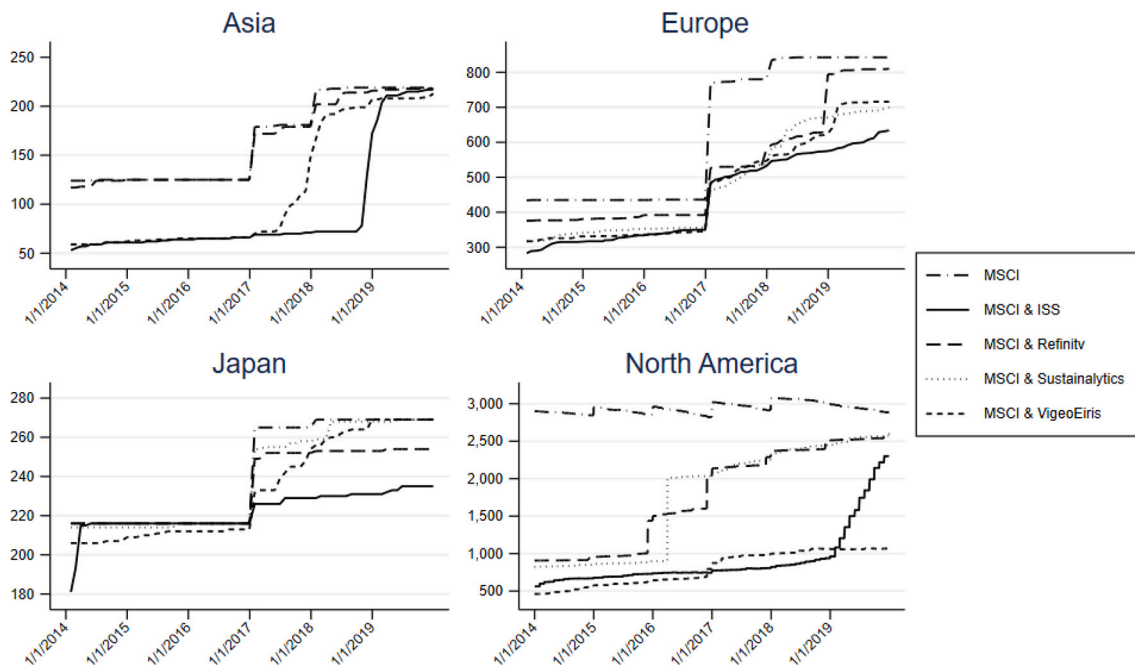


Fig. A2. Number of stocks that are rated by MSCI in total and number of stocks considered in rating pairs in the period from January 2014 to December 2019.

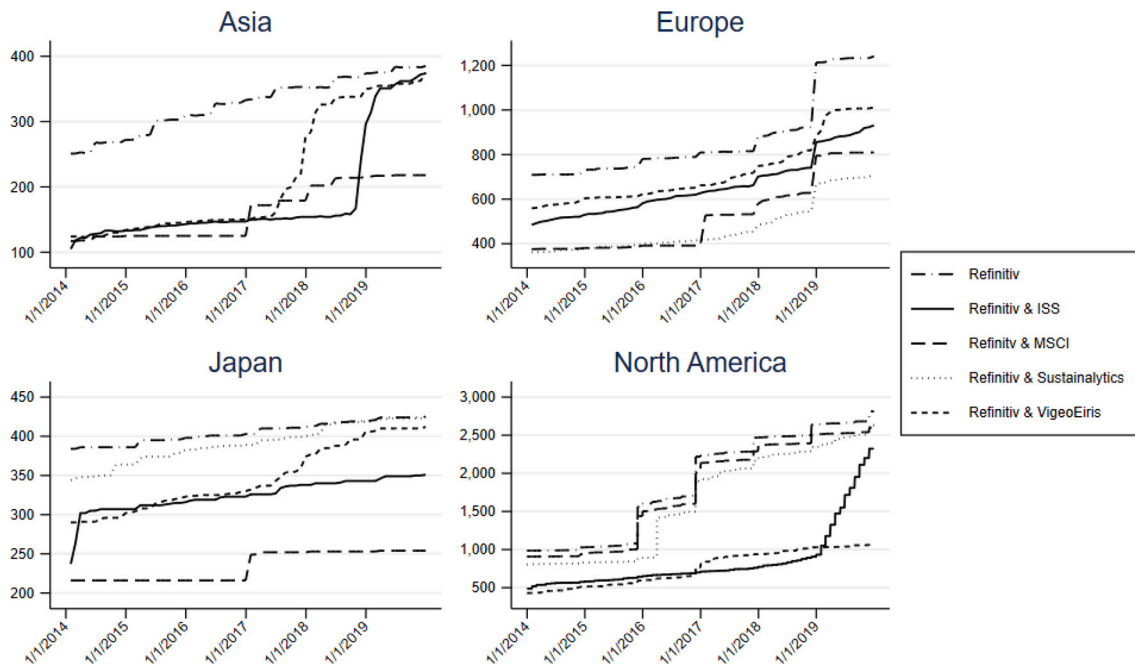


Fig. A3. Number of stocks that are rated by Refinitiv in total and number of stocks considered in rating pairs in the period from January 2014 to December 2019.

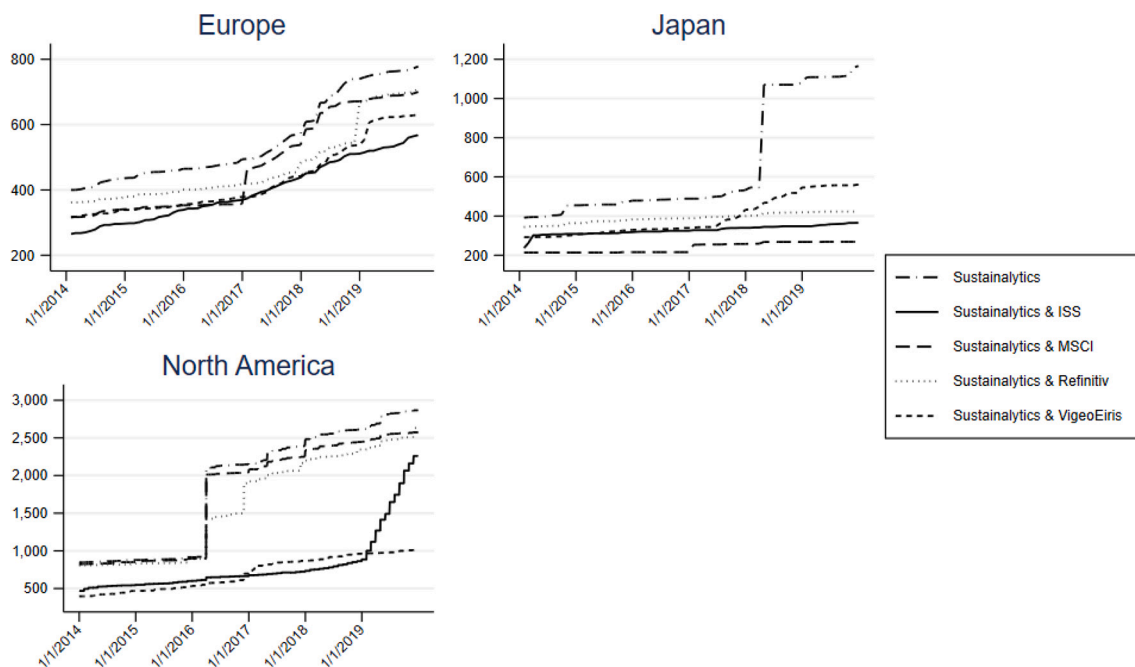


Fig. A4. Number of stocks that are rated by Sustainalytics in total and number of stocks considered in rating pairs in the period from January 2014 to December 2019.

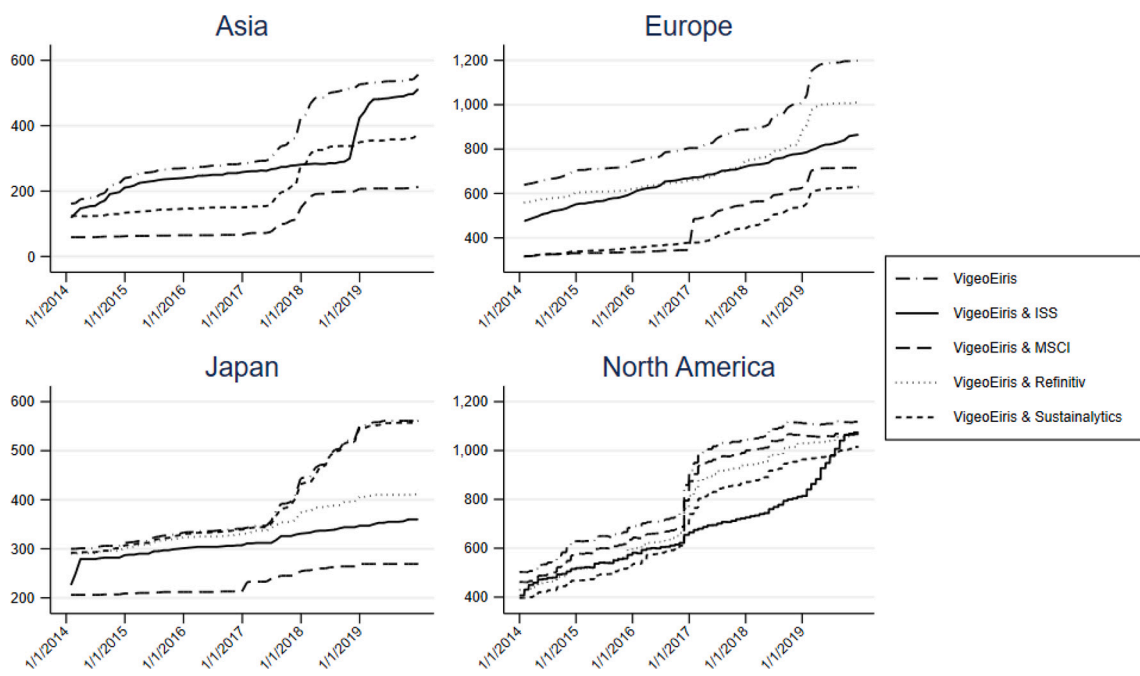


Fig. A5. Number of stocks that are rated by VigeoEiris in total and number of stocks considered in rating pairs in the period from January 2014 to December 2019.



Fig. A6. Industry exposure (market capitalization weighted) of North American portfolios. Notes: We present plots of the mean industry exposure of North American quintile portfolios weighted by stocks' market capitalization.

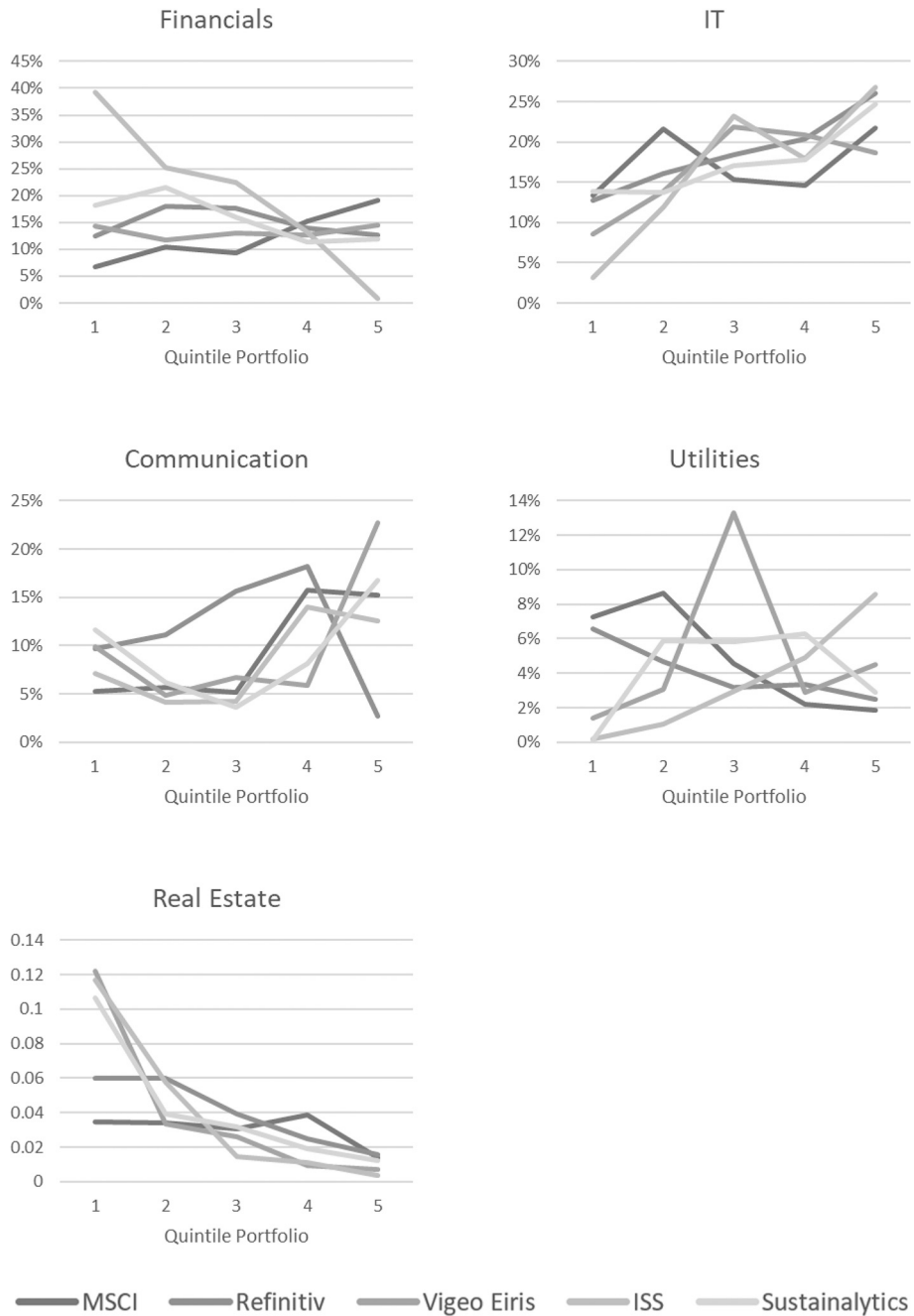


Fig. A7. Industry exposure (market capitalization weighted) of North American portfolios (cont'd).

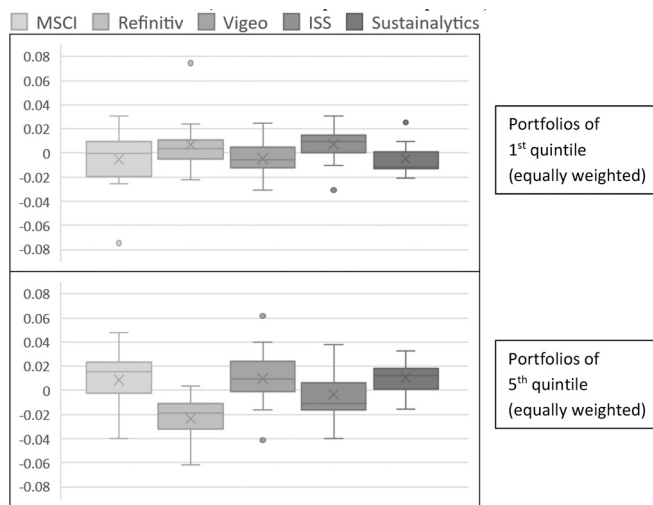


Fig. A8. Differences in daily idiosyncratic risk between portfolios in the period from January 2014 to December 2019 (all values are presented in percent).

Table A1  
Quintile portfolios' return.

Quintile	Return		
	Mean [min; max]		
	Equally weighted	Mcap weighted	Min variance optimization
1	0.97 [0.83; 1.17]	1.24 [0.94; 1.74]	0.56 [ 2.74; 3.09]
2	0.90 [0.56; 1.36]	1.09 [0.55; 1.76]	0.93 [ 1.70; 4.96]
3	0.81 [0.60; 1.12]	0.92 [0.75; 1.31]	0.58 [ 1.67; 2.62]
4	0.75 [0.61; 0.97]	0.86 [0.64; 1.23]	1.41 [ 0.29; 2.77]
5	0.62 [0.45; 0.74]	0.49 [0.40; 0.61]	0.07 [ 2.94; 3.30]
5-1	0.35	0.76	0.48
≤1 rating	[ 0.56; 0.12] 0.63	[ 1.28; 0.44] 1.52	[ 3.83; 3.91] 1.69
1	0.51 [0.36; 0.67]	0.94 [0.68; 1.13]	0.38 [ 5.77; 4.60]
2	0.58 [0.35; 0.79]	0.94 [0.77; 1.15]	0.64 [ 3.31; 4.73]
3	0.45 [0.34; 0.58]	0.60 [0.41; 1.04]	0.20 [ 4.52; 3.63]
4	0.49 [0.35; 0.57]	0.64 [0.48; 0.79]	0.40 [ 2.97; 4.90]
5	0.47 [0.38; 0.61]	0.67 [0.54; 0.81]	0.00 [ 3.23; 3.12]
5-1	0.03	0.27	0.38
≤1 rating	[ 0.29; 0.09] 0.03	[ 0.45; 0.13] 1.22	[ 6.24; 6.59] 0.55
1	0.66 [0.53; 0.83]	1.12 [0.99; 1.32]	1.86 [ 5.47; 3.16]
2	0.64 [0.55; 0.80]	0.85 [0.67; 1.10]	0.91 [ 2.27; 5.07]
3	0.78 [0.64; 0.90]	1.03 [0.83; 1.33]	0.82 [ 3.34; 3.92]
4	0.74 [0.66; 0.81]	0.94 [0.82; 1.10]	2.74 [ 1.28; 7.71]
5	0.76 [0.67; 0.88]	0.91 [0.79; 1.04]	1.50 [ 3.93; 0.81]
5-1	0.10	0.21	0.36
≤1 rating	[ 0.14; 0.31] 0.74	[ 0.48; 0.02] 1.53	[ 3.80; 5.39] 1.64

(continued on next page)

Table A1 (continued)

Quintile	Return		
	Mean		
	[min; max]		
	Equally weighted	Mcap weighted	Min variance optimization
1	0.75 [0.55; 1.03]	1.42 [1.26; 1.66]	0.52 [ 3.10; 4.93]
2	0.77 [0.62; 0.96]	1.26 [0.98; 1.56]	1.99 [ 1.58; 6.72]
3	0.75 [0.63; 0.90]	1.21 [1.08; 1.37]	0.51 [ 4.02; 5.14]
4	0.80 [0.68; 0.95]	1.17 [1.03; 1.31]	0.84 [ 2.42; 3.17]
5	0.87 [0.77; 0.99]	1.21 [1.12; 1.29]	1.36 [ 0.56; 6.18]
5-1	0.12 [ 0.12; 0.35]	0.21 [ 0.39; 0.02]	0.84 [ 4.52; 8.32]
≤1 rating	0.07	1.49	1.00

Notes: The table provides the mean, minimum, and maximum of the monthly return of the quintile portfolios of the 36 rating pairs and the portfolio of stocks with only one or no ESG rating (≤1 rating) per geographic region in the period from January 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. We compute the mean monthly return for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we take the latter values and compute their mean, minimum and maximum values. The minimum and maximum values are shown in the squared brackets. In addition, we compute differences between the fifth quintile portfolios and the first quintile portfolios of all rating pairs. We report the respective mean, minimum, and maximum values of these differences in the lines "5-1". Example: When portfolio returns are computed with equally weighted returns, on average, the mean monthly returns of the first quintile portfolios covering North American stocks is 0.97 %, with a minimum value of 0.83 % and a maximum value of 1.17 %. The difference between the mean return of the fifth quintile portfolios and the first quintile portfolios is 0.53 % on average, with a minimum value of 0.56 % and a maximum value of 0.12 %.

Table A2  
Quintile portfolios' alpha and idiosyncratic risk (FF-3 factor model).

Quintile	Alpha (FF-3)			Idiosyncratic risk (FF-3)		
	Mean			Mean		
	[min; max]			[min; max]		
	Equally weighted	Mcap weighted	Min variance optimization	Equally weighted	Mcap weighted	Min variance optimization
<i>Panel A: Asia-Pacific</i>						
1	0.77 [0.39; 1.24]	1.03 [0.58; 1.96]	0.52 [ 1.52; 2.52]	0.50 [0.35; 0.63]	0.57 [0.46; 0.66]	0.57 [0.46; 0.66]
2	0.58 [0.15; 1.41]	0.76 [0.28; 1.68]	0.09 [ 1.98; 1.85]	0.51 [0.32; 0.87]	0.56 [0.47; 0.85]	0.56 [0.47; 0.85]
3	0.47 [0.19; 0.73]	0.61 [0.36; 1.10]	0.46 [ 0.97; 1.53]	0.46 [0.34; 0.59]	0.51 [0.38; 0.57]	0.51 [0.38; 0.57]
4	0.32 [ 0.09; 0.54]	0.53 [ 0.02; 0.92]	0.01 [ 0.80; 1.54]	0.46 [0.30; 0.64]	0.48 [0.36; 0.56]	0.48 [0.36; 0.56]
5	0.13 [ 0.14; 0.32]	0.16 [ 0.04; 0.35]	0.48 [ 2.00; 0.49]	0.45 [0.34; 0.57]	0.43 [0.31; 0.55]	0.43 [0.31; 0.55]
5-1	0.64 [ 1.07; 0.16]	0.88 [ 1.83; 0.28]	1.00 [ 3.23; 1.59]	0.05 [ 0.28; 0.04]	0.14 [ 0.34; 0.02]	0.14 [ 0.34; 0.02]
≤1 rating	0.70	1.54	1.67	0.37	0.48	0.34
<i>Panel B: Europe</i>						
1	0.18 [0.02; 0.33]	0.58 [0.22; 0.75]	0.19 [ 1.57; 2.91]	0.35 [0.33; 0.37]	0.35 [0.33; 0.41]	0.35 [0.33; 0.41]
2	0.23 [ 0.02; 0.40]	0.55 [0.29; 0.73]	1.15 [ 0.75; 3.44]	0.35 [0.33; 0.38]	0.35 [0.33; 0.39]	0.35 [0.33; 0.39]
3	0.19 [ 0.01; 0.34]	0.40 [0.18; 0.67]	0.15 [ 3.48; 1.08]	0.35 [0.32; 0.38]	0.33 [0.30; 0.36]	0.33 [0.30; 0.36]
4	0.19 [0.06; 0.32]	0.40 [0.23; 0.55]	0.27 [ 1.99; 1.34]	0.33 [0.32; 0.34]	0.31 [0.29; 0.33]	0.31 [0.29; 0.33]
5	0.22 [0.08; 0.38]	0.49 [0.40; 0.64]	0.85 [ 2.05; 0.95]	0.32 [0.30; 0.34]	0.29 [0.27; 0.30]	0.29 [0.27; 0.30]
5-1	0.04 [ 0.18; 0.23]	0.08 [ 0.25; 0.22]	1.04 [ 4.15; 0.88]	0.03 [ 0.06; 0.01]	0.07 [ 0.13; 0.04]	0.07 [ 0.13; 0.04]
≤1 rating	0.20	1.17	0.44	0.25	0.34	0.63

(continued on next page)

**Table A2** (continued)

Quintile	Alpha (FF-3)			Idiosyncratic risk (FF-3)		
	Mean [min; max]			Mean [min; max]		
	Equally weighted	Mcap weighted	Min variance optimization	Equally weighted	Mcap weighted	Min variance optimization
<i>Panel C: Japan</i>						
1	0.25 [0.02; 0.52]	0.60 [0.42; 0.88]	0.76 [ 0.53; 3.15]	0.38 [0.36; 0.40]	0.39 [0.38; 0.41]	0.39 [0.38; 0.41]
2	0.29 [0.17; 0.49]	0.48 [0.27; 0.80]	1.14 [ 0.37; 2.46]	0.35 [0.34; 0.36]	0.37 [0.35; 0.40]	0.37 [0.35; 0.40]
3	0.47 [0.33; 0.66]	0.71 [0.48; 1.08]	1.24 [0.07; 2.95]	0.34 [0.33; 0.35]	0.36 [0.34; 0.39]	0.36 [0.34; 0.39]
4	0.44 [0.36; 0.54]	0.70 [0.46; 0.88]	0.31 [ 1.45; 1.60]	0.34 [0.34; 0.36]	0.34 [0.33; 0.35]	0.34 [0.33; 0.35]
5	0.45 [0.36; 0.57]	0.71 [0.52; 0.95]	0.09 [ 0.97; 0.91]	0.33 [0.32; 0.35]	0.34 [0.31; 0.38]	0.34 [0.31; 0.38]
5-1	0.20 [ 0.06; 0.50]	0.11 [ 0.07; 0.34]	0.86 [ 3.53; 1.00]	0.05 [ 0.08; 0.02]	0.06 [ 0.10; 0.01]	0.06 [ 0.10; 0.01]
≤1 rating	0.16	0.96	1.06	0.27	0.31	0.43
<i>Panel D: North America</i>						
1	0.23 [ 0.01; 0.50]	0.60 [0.40; 0.88]	0.69 [ 1.53; 3.38]	0.18 [0.16; 0.22]	0.21 [0.19; 0.25]	0.21 [0.19; 0.25]
2	0.08 [ 0.16; 0.47]	0.40 [0.13; 0.91]	0.42 [ 2.01; 2.13]	0.16 [0.13; 0.23]	0.19 [0.17; 0.25]	0.19 [0.17; 0.25]
3	0.06 [ 0.08; 0.28]	0.31 [0.12; 0.54]	0.16 [ 1.57; 1.53]	0.15 [0.13; 0.16]	0.17 [0.15; 0.19]	0.17 [0.15; 0.19]
4	0.11 [ 0.03; 0.28]	0.30 [0.17; 0.43]	0.70 [ 1.43; 1.82]	0.15 [0.11; 0.17]	0.14 [0.12; 0.16]	0.14 [0.12; 0.16]
5	0.21 [0.08; 0.29]	0.30 [0.21; 0.36]	0.60 [ 1.18; 3.07]	0.15 [0.10; 0.21]	0.11 [0.07; 0.14]	0.11 [0.07; 0.14]
5-1	0.02 [ 0.42; 0.25]	0.30 [ 0.62; 0.07]	0.08 [ 2.57; 4.59]	0.03 [ 0.08; 0.04]	0.10 [ 0.16; 0.05]	0.10 [ 0.16; 0.05]
≤1 rating	0.16	1.01	1.00	0.15	0.17	0.72

*Notes:* The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly alpha and daily idiosyncratic risk per quintile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). Portfolio alpha is computed by running Eq. (3) without the factors RMW and CMA on the daily portfolio returns of every month. We multiply the respective alpha with the number of observations per month to compute the monthly alpha. We compute the mean monthly alpha for every quintile portfolio of the 36 rating pairs per geographic region. Idiosyncratic risk per portfolio is computed by first running Eq. (3) without the factors RMW and CMA on the daily portfolio returns of every month to estimate the betas in Eq. (3). Thereafter, the betas are used to compute the residual  $\epsilon$  for every observation. Idiosyncratic risk is then computed as the standard deviation of the residuals of every month. We compute the mean daily idiosyncratic risk per month for every quintile portfolio of the 36 rating pairs per geographic region. The table shows the mean of the absolute values. The minimum and maximum values are shown in the squared brackets. Example: When portfolio returns are computed with equally weighted returns, on average, the mean monthly alpha of the first quintile portfolio of a rating provider is 0.23 % for North American stocks, with a minimum value of 0.01 % and a maximum value of 0.50 %.

**Table A3**

Intersection of quintile portfolios in the period from December 2014 to December 2019.

<i>Panel A: Asia-Pacific</i>					
Quintile provider A	Quintile provider B				
	1	2	3	4	5
1	49.1 [40.8; 58.7]	27.7 [23.2; 31.1]	14.3 [11.2; 18.8]	6.2 [4.0; 9.3]	2.7 [0.5; 6.1]
2	26.3 [20.7; 28.9]	27.7 [17.4; 36.4]	23.7 [16.4; 29.1]	16.4 [12.4; 24.4]	5.9 [3.5; 12.2]
3	17.2 [10.1; 22.2]	25.7 [22.3; 29.2]	26.0 [16.5; 33.5]	22.2 [21.1; 23.6]	8.9 [5.6; 15.0]
4	5.9 [3.2; 7.8]	15.1 [11.5; 20.1]	25.0 [22.7; 28.0]	28.0 [22.2; 34.2]	25.9 [23.0; 28.5]
5	0.9 [0.0; 2.3]	4.5 [1.5; 9.5]	10.5 [4.9; 17.7]	28.4 [25.7; 32.5]	55.6 [39.5; 65.5]
<i>Panel B: Europe</i>					
1	53.6 [40.6; 62.3]	26.0 [22.5; 28.1]	11.9 [7.5; 18.5]	5.5 [2.3; 10.7]	3.1 [0.3; 8.1]
2	28.6 [24.5; 30.8]	30.8 [26.9; 35.8]	22.9 [19.7; 27.9]	12.5 [9.4; 16.8]	5.2 [2.5; 7.9]
3	13.2 [7.5; 18.2]	24.2 [22.0; 29.4]	28.5 [23.9; 33.5]	22.5 [17.3; 27.0]	11.6 [10.0; 14.2]

(continued on next page)

**Table A3** (continued)

Panel A: Asia-Pacific					
Quintile provider A	Quintile provider B				
	1	2	3	4	5
4	5.1 [1.7; 8.7]	14.4 [10.7; 17.7]	22.9 [21.5; 25.7]	29.9 [26.5; 34.3]	27.7 [25.1; 30.1]
5	1.4 [0.2; 3.1]	5.0 [1.9; 9.5]	13.3 [9.9; 17.4]	29.8 [26.9; 31.9]	50.4 [44.3; 57.9]
Panel C: Japan					
1	58.9 [45.7; 72.3]	22.3 [16.8; 28.1]	10.6 [6.3; 15.5]	5.6 [0.5; 10.9]	2.7 [0.1; 4.9]
2	24.7 [20.3; 29.2]	32.2 [23.3; 39.4]	23.1 [20.3; 25.5]	13.4 [7.3; 22.0]	6.6 [1.4; 12.1]
3	11.6 [6.6; 19.4]	24.4 [20.8; 27.6]	28.5 [23.5; 34.5]	21.4 [16.2; 26.0]	14.0 [8.6; 19.9]
4	4.7 [1.5; 8.4]	13.9 [8.2; 20.6]	23.4 [20.5; 25.9]	29.5 [26.0; 36.2]	28.5 [22.4; 32.0]
5	1.8 [0.3; 4.0]	6.9 [3.5; 10.9]	14.7 [9.3; 18.6]	29.8 [25.6; 33.4]	46.8 [37.9; 56.7]
Panel D: North America					
1	45.4 [29.7; 59.8]	28.0 [25.3; 30.8]	15.7 [11.4; 20.3]	7.8 [2.4; 16.2]	3.1 [0.1; 8.4]
2	29.1 [23.7; 37.3]	28.9 [22.5; 33.8]	23.3 [19.3; 26.1]	14.0 [9.8; 20.5]	4.7 [1.9; 8.3]
3	17.1 [12.0; 24.0]	24.0 [21.1; 28.4]	26.2 [18.7; 33.1]	22.4 [19.9; 24.8]	10.2 [6.5; 15.0]
4	8.2 [1.8; 16.6]	14.2 [9.6; 18.5]	21.9 [18.3; 24.7]	30.4 [24.7; 36.7]	25.3 [20.5; 29.1]
5	2.7 [0.2; 6.0]	5.0 [1.2; 9.7]	10.9 [6.3; 13.7]	25.5 [18.0; 29.2]	55.9 [44.0; 66.7]

Notes: The table provides the mean, minimum, and maximum pairwise intersection of the portfolios of the rating pairs subdivided by geographic region. The minimum and maximum values of the size of the intersection are shown in the squared brackets. All values are presented in percent. Example: In the sample of North American stocks, on average, 28.9 % of the stocks included in the second quintile portfolio of rating provider A are also in the second quintile portfolio of rating provider B. The minimum (maximum) values of the size of the intersection for the second quintile portfolios is 22.5 % (33.8 %).

**Table A4**  
Differences between quintile portfolios' return.

Quintile	Return					
	Equally weighted		Mcap weighted		Min variance optimization	
	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value
Panel A: Asia						
1	0.09 [0.00; 0.19]	81.9	0.21 [0.03; 0.49]	52.6	0.71 [0.25; 1.13]	43.5
2	0.18 [0.07; 0.25]	72.8	0.22 [0.00; 0.47]	54.9	1.21 [0.09; 2.29]	14.1
3	0.11 [0.00; 0.27]	74.6	0.12 [0.00; 0.36]	63.1	0.83 [0.14; 1.61]	22.8
4	0.14 [0.00; 0.36]	65.5	0.21 [0.02; 0.41]	58.3	0.79 [0.22; 1.34]	28.6
5	0.08 [0.03; 0.13]	86.6	0.05 [0.01; 0.09]	89.0	0.46 [0.10; 1.18]	24.0
Panel B: Europe						
1	0.08 [0.00; 0.26]	71.7	0.17 [0.09; 0.39]	56.5	1.45 [0.03; 2.53]	15.3
2	0.14 [0.01; 0.32]	64.1	0.15 [0.05; 0.25]	70.4	1.11 [0.08; 2.01]	20.0
3	0.06 [0.01; 0.14]	84.3	0.21 [0.01; 0.60]	36.3	1.32 [0.43; 2.90]	3.8
4	0.08 [0.00; 0.20]	76.3	0.11 [0.04; 0.24]	72.0	0.91 [0.07; 2.54]	6.2
5	0.08 [0.00; 0.17]	80.1	0.11 [0.03; 0.20]	75.0	0.96 [0.06; 2.27]	4.4

Panel C: Japan

(continued on next page)

**Table A4** (continued)

Quintile	Return					
	Equally weighted		Mcap weighted		Min variance optimization	
	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value
1	0.07 [0.03; 0.10]	85.4	0.13 [0.02; 0.31]	60.2	0.77 [0.09; 1.44]	23.0
2	0.08 [0.01; 0.25]	66.3	0.10 [0.01; 0.31]	57.7	0.76 [0.05; 1.63]	24.2
3	0.10 [0.00; 0.22]	71.3	0.19 [0.03; 0.32]	60.4	0.90 [0.19; 2.60]	2.4
4	0.06 [0.00; 0.13]	84.3	0.09 [0.00; 0.26]	66.2	0.81 [0.19; 2.08]	5.1
5	0.08 [0.00; 0.18]	76.0	0.10 [0.02; 0.17]	75.9	0.51 [0.19; 0.77]	49.1
<i>Panel D: North America</i>						
1	0.11 [0.02; 0.28]	64.9	0.13 [0.01; 0.35]	58.1	1.41 [0.03; 2.77]	6.9
2	0.11 [0.00; 0.25]	70.6	0.20 [0.00; 0.58]	42.4	1.04 [0.04; 2.41]	13.4
3	0.07 [0.01; 0.13]	83.7	0.10 [0.02; 0.25]	67.3	0.55 [0.00; 1.73]	32.3
4	0.08 [0.01; 0.18]	77.3	0.09 [0.02; 0.21]	70.8	0.80 [0.21; 1.36]	25.1
5	0.08 [0.01; 0.21]	71.7	0.08 [0.01; 0.16]	75.7	1.12 [0.12; 2.26]	14.0

*Notes:* The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly return per quintile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted") and weights derived with minimum variance optimization (see columns "Min variance optimization"). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. We compute the mean monthly return for every quintile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between the mean returns of the two matching quintile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum p-value of the t-tests per two matching quintile portfolios among the rating pairs. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean monthly return between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 0.11 % for North American stocks, with a minimum value of 0.02 % and a maximum value of 0.28 %. The smallest p-value of the t-tests on the difference in return between the first quintile portfolio of a rating provider A and the first quintile portfolio of a provider B is 64.9 % for North American stocks. This shows that the monthly returns of the two most different portfolios are not different from each other at statistically significant levels.

**Table A5**

Intersection of quartile portfolios of 36 rating pairs.

Quintile provider A	Quintile provider B			
	1	2	3	4
1	57.5 [37.3; 75.7]	26.6 [19.3; 33.2]	11.6 [4.1; 23.8]	4.3 [0.2; 10.5]
2	28.7 [20.9; 37.9]	35.1 [22.4; 48.5]	25.9 [21.2; 31.0]	10.3 [5.5; 20.0]
3	12.4 [4.6; 25.5]	27.0 [22.4; 32.7]	33.9 [26.8; 41.1]	26.7 [19.0; 33.1]
4	3.1 [0.1; 8.9]	10.9 [4.7; 20.1]	28.7 [19.3; 35.2]	57.3 [42.8; 71.6]

*Notes:* The table provides the mean, minimum, and maximum pairwise stock overlap of the 288 portfolios of the 36 rating pairs in the period from January 2014 to December 2019. The minimum and maximum stock overlap are shown in the squared brackets. All values are presented in percent. Example: On average, 35.1 % of the stocks included in the second quartile portfolio of rating provider A are also in the second quartile portfolio of rating provider B. The minimum (maximum) stock overlap for the second quartile portfolios is 22.4 % (48.5 %).

**Table A6**  
Intersection of sextile portfolios of 36 rating pairs.

Quintile provider A	Quintile provider B					
	1	2	3	4	5	6
1	48.5 [27.6; 66.7]	24.4 [15.1; 31.0]	13.4 [6.1; 21.6]	7.7 [1.8; 16.8]	4.1 [0.0; 11.4]	1.9 [0.0; 8.1]
2	26.4 [18.2; 34.7]	26.8 [17.7; 36.6]	21.4 [16.1; 25.5]	13.9 [6.2; 24.8]	7.6 [1.8; 15.9]	3.8 [0.6; 8.9]
3	14.9 [6.1; 25.1]	22.0 [15.0; 26.6]	23.1 [13.2; 32.6]	20.4 [16.2; 28.9]	13.6 [9.9; 20.8]	6.1 [2.2; 12.5]
4	8.2 [1.3; 20.6]	14.6 [9.6; 21.6]	21.1 [14.9; 27.4]	22.8 [15.1; 28.8]	20.3 [13.5; 28.2]	13.1 [6.6; 21.7]
5	3.3 [0.4; 14.1]	7.6 [3.2; 13.7]	14.3 [9.7; 20.7]	21.1 [16.2; 26.5]	27.7 [19.9; 34.6]	26.0 [20.9; 35.4]
6	1.1 [0.0; 5.2]	3.1 [0.3; 9.2]	6.9 [2.0; 15.1]	14.3 [7.5; 21.8]	27.4 [20.0; 33.4]	47.2 [32.7; 63.1]

Notes: The table provides the mean, minimum, and maximum pairwise stock overlap of the 432 portfolios of the 36 rating pairs in the period from January 2014 to December 2019. The minimum and maximum stock overlap are shown in the squared brackets. All values are presented in percent. Example: On average, 26.8 % of the stocks included in the second sextile portfolio of rating provider A are also in the second sextile portfolio of rating provider B. The minimum (maximum) stock overlap for the second sextile portfolios is 17.7 % (36.6 %).

**Table A7**  
Differences between quartile portfolios' return and alpha.

Quartile	Return				Alpha			
	Equally weighted		Mcap weighted		Equally weighted		Mcap weighted	
	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value
<i>Panel A: Asia</i>								
1	0.07 [0.03; 0.11]	89.6	0.20 [0.09; 0.26]	75.2	0.13 [0.01; 0.24]	59.1	0.21 [0.02; 0.46]	33.8
2	0.18 [0.01; 0.30]	72.1	0.21 [0.01; 0.39]	61.7	0.16 [0.01; 0.44]	38.4	0.26 [0.03; 0.47]	29.7
3	0.13 [0.03; 0.24]	73.5	0.17 [0.07; 0.31]	69.1	0.19 [0.06; 0.35]	21.6	0.19 [0.07; 0.44]	22.7
4	0.10 [0.00; 0.28]	70.8	0.05 [0.00; 0.11]	87.0	0.26 [0.07; 0.47]	21.3	0.13 [0.02; 0.31]	27.5
<i>Panel B: Europe</i>								
1	0.05 [0.00; 0.10]	88.7	0.11 [0.01; 0.28]	65.5	0.05 [0.01; 0.12]	62.8	0.15 [0.01; 0.33]	15.1
2	0.10 [0.01; 0.21]	76.1	0.13 [0.03; 0.23]	72.8	0.11 [0.01; 0.32]	23.4	0.10 [0.02; 0.23]	40.3
3	0.07 [0.02; 0.16]	82.1	0.10 [0.01; 0.18]	78.8	0.08 [0.02; 0.15]	51.5	0.05 [0.00; 0.10]	63.8
4	0.08 [0.01; 0.13]	83.9	0.08 [0.00; 0.18]	76.7	0.06 [0.01; 0.14]	50.8	0.04 [0.01; 0.07]	70.3
<i>Panel C: Japan</i>								
1	0.07 [0.01; 0.14]	79.3	0.12 [0.02; 0.26]	66.3	0.12 [0.01; 0.27]	29.9	0.15 [0.01; 0.30]	29.0
2	0.09 [0.01; 0.16]	77.8	0.12 [0.03; 0.33]	56.8	0.08 [0.00; 0.14]	46.8	0.12 [0.01; 0.25]	32.1
3	0.05 [0.00; 0.14]	82.7	0.05 [0.00; 0.11]	85.5	0.14 [0.06; 0.22]	28.2	0.11 [0.01; 0.22]	35.0
4	0.08 [0.01; 0.15]	80.2	0.09 [0.03; 0.20]	73.2	0.08 [0.01; 0.17]	36.7	0.07 [0.00; 0.15]	41.2
<i>Panel D: North America</i>								
1	0.11 [0.01; 0.23]	71.6	0.14 [0.03; 0.29]	64.1	0.10 [0.01; 0.21]	17.3	0.23 [0.00; 0.53]	1.5
2	0.09 [0.04; 0.16]	80.0	0.13 [0.01; 0.26]	66.2	0.10 [0.01; 0.20]	18.8	0.09 [0.01; 0.21]	18.6
3	0.07 [0.02; 0.13]	85.7	0.05 [0.01; 0.15]	78.9	0.10 [0.01; 0.23]	9.6	0.08 [0.00; 0.20]	15.2
4	0.08 [0.00; 0.19]	74.4	0.07 [0.01; 0.15]	77.5	0.04 [0.00; 0.11]	39.9	0.04 [0.00; 0.06]	56.0

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly return and alpha per quartile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted" for the latter). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. Portfolio alpha is computed by running Eq. (3) on the

daily portfolio returns of every month. We multiply the respective alpha with the number of observations per month to compute the monthly alpha. We compute the mean monthly return and alpha for every quartile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between mean return/alpha of the two matching quartile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum *p*-value of the *t*-tests per two matching quartile portfolios among the rating pairs. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean monthly return between the first quartile portfolio of a rating provider A and the first quartile portfolio of a provider B is 0.11 % for North American stocks, with a minimum value of 0.01 % and a maximum value of 0.23 %. The smallest *p*-value of the *t*-tests on the difference in return between the first quartile portfolio of a rating provider A and the first quartile portfolio of a provider B is 71.6 % for North American stocks. This shows that the monthly returns of the two most different portfolios are not different from each other at statistically significant levels.

**Table A8**  
Differences between sextile portfolios' return and alpha.

Sextile	Return				Alpha			
	Equally weighted		Mcap weighted		Equally weighted		Mcap weighted	
	Mean [min; max] of mean difference]	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value
<i>Panel A: Asia</i>								
1	0.12 [0.06; 0.25]	77.6	0.29 [0.07; 0.54]	52.4	0.15 [0.04; 0.31]	52.5	0.22 [0.01; 0.73]	17.3
2	0.14 [0.06; 0.27]	75.8	0.35 [0.18; 0.52]	55.6	0.20 [0.03; 0.67]	16.3	0.46 [0.11; 0.66]	17.6
3	0.21 [0.07; 0.38]	59.2	0.24 [0.01; 0.40]	60.5	0.22 [0.06; 0.35]	28.6	0.30 [0.21; 0.43]	30.0
4	0.16 [0.04; 0.37]	63.2	0.20 [0.02; 0.55]	50.9	0.24 [0.03; 0.50]	14.4	0.29 [0.14; 0.48]	25.3
5	0.19 [0.04; 0.31]	70.3	0.24 [0.00; 0.45]	55.3	0.25 [0.01; 0.78]	10.1	0.09 [0.01; 0.26]	35.7
6	0.14 [0.04; 0.19]	78.9	0.09 [0.03; 0.19]	78.7	0.32 [0.15; 0.53]	21.9	0.15 [0.07; 0.32]	31.2
<i>Panel B: Europe</i>								
1	0.09 [0.01; 0.16]	81.1	0.13 [0.01; 0.37]	58.7	0.14 [0.03; 0.27]	36.9	0.21 [0.02; 0.57]	5.8
2	0.13 [0.03; 0.26]	71.5	0.09 [0.01; 0.15]	81.9	0.16 [0.01; 0.43]	13.0	0.21 [0.06; 0.33]	23.8
3	0.13 [0.01; 0.24]	72.9	0.14 [0.02; 0.34]	62.9	0.16 [0.02; 0.30]	31.5	0.16 [0.03; 0.29]	31.6
4	0.08 [0.01; 0.22]	75.5	0.13 [0.01; 0.29]	65.3	0.09 [0.00; 0.21]	34.3	0.12 [0.00; 0.27]	22.0
5	0.12 [0.01; 0.30]	65.9	0.08 [0.00; 0.23]	71.5	0.10 [0.01; 0.22]	31.1	0.07 [0.02; 0.15]	52.4
6	0.08 [0.01; 0.16]	81.5	0.15 [0.03; 0.26]	68.2	0.06 [0.00; 0.15]	43.6	0.11 [0.02; 0.20]	27.8
<i>Panel C: Japan</i>								
1	0.05 [0.00; 0.15]	78.4	0.11 [0.03; 0.25]	66.8	0.14 [0.03; 0.29]	32.8	0.18 [0.04; 0.36]	23.0
2	0.07 [0.01; 0.11]	84.1	0.12 [0.01; 0.22]	68.4	0.06 [0.01; 0.15]	55.0	0.11 [0.01; 0.28]	37.6
3	0.15 [0.01; 0.29]	63.1	0.14 [0.02; 0.33]	57.5	0.13 [0.00; 0.33]	16.9	0.17 [0.02; 0.30]	30.9
4	0.10 [0.01; 0.20]	74.2	0.08 [0.02; 0.15]	79.6	0.17 [0.00; 0.35]	15.6	0.11 [0.00; 0.19]	42.9
5	0.06 [0.00; 0.23]	71.0	0.17 [0.06; 0.42]	46.9	0.07 [0.01; 0.22]	34.5	0.11 [0.00; 0.28]	22.8
6	0.07 [0.00; 0.16]	80.1	0.12 [0.05; 0.22]	68.8	0.07 [0.01; 0.15]	47.1	0.14 [0.02; 0.26]	23.7
<i>Panel D: North America</i>								
1	0.12 [0.02; 0.32]	60.7	0.13 [0.00; 0.34]	59.1	0.08 [0.01; 0.28]	15.8	0.24 [0.05; 0.47]	5.1
2	0.13 [0.02; 0.41]	66.5	0.27 [0.03; 0.89]	29.1	0.14 [0.02; 0.46]	11.3	0.20 [0.00; 0.56]	15.7
3	0.09 [0.01; 0.19]	77.4	0.13 [0.00; 0.29]	62.6	0.10 [0.00; 0.25]	9.2	0.10 [0.00; 0.24]	23.2
4	0.10 [0.01; 0.19]	77.5	0.04 [0.00; 0.11]	85.5	0.16 [0.00; 0.29]	5.8	0.13 [0.00; 0.36]	3.9
5	0.04 [0.01; 0.10]	87.7	0.06 [0.01; 0.13]	81.3	0.08 [0.00; 0.27]	6.6	0.10 [0.01; 0.24]	7.2
6	0.08 [0.01; 0.24]	68.4	0.08 [0.00; 0.18]	71.9	0.07 [0.01; 0.15]	19.3	0.03 [0.00; 0.07]	57.3

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in monthly return and alpha per sextile portfolio and geographic region of the rating pairs in the period from December 2014 to December 2019. All values are presented in percent. Portfolio returns are computed with equally weighted returns as well as with returns weighted by stocks' market capitalization (see columns "Mcap weighted" for the latter). We multiply the mean value of the daily portfolio returns per month with the number of observations per month to compute the monthly returns. Portfolio alpha is computed by running Eq. (3) on the daily portfolio returns of every month. We multiply the respective alpha with the number of observations per month to compute the monthly alpha. We compute the mean monthly return and alpha for every sextile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between mean return/alpha of the two matching sextile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum *p*-value of the *t*-tests per two matching sextile portfolios among the rating pairs. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean monthly return between the first sextile portfolio of a rating provider A and the first sextile portfolio of a provider B is 0.12 % for North American stocks, with a minimum value of 0.02 % and a maximum value of 0.32 %. The smallest *p*-value of the *t*-tests on the difference in return between the first sextile portfolio of a rating provider A and the first sextile portfolio of a provider B is 60.7 % for North American stocks. This shows that the monthly returns of the two most different portfolios are not different from each other at statistically significant levels.

**Table A9**  
Differences between quartile and sextile portfolios' idiosyncratic risk.

Quartile/Sextile	Idiosyncratic risk quartile portfolios				Idiosyncratic risk sextile portfolios			
	Equally weighted		Mcap weighted		Equally weighted		Mcap weighted	
	Mean [min; max of mean difference]	T-test min p-value	Mean [min; max of mean difference]	T-test min p-value	Mean [min; max of mean difference]	T-test min p-value	Mean [min; max of mean difference]	T-test min p-value
<i>Panel A: Asia</i>								
1	0.02 [0.00; 0.05]	2.5	0.05 [0.02; 0.10]	0.0	0.03 [0.00; 0.10]	0.0	0.06 [0.00; 0.15]	0.0
2	0.04 [0.00; 0.06]	0.2	0.03 [0.01; 0.06]	1.6	0.04 [0.02; 0.06]	4.1	0.04 [0.01; 0.16]	0.0
3	0.03 [0.00; 0.06]	0.1	0.06 [0.01; 0.14]	0.0	0.03 [0.01; 0.05]	2.8	0.04 [0.00; 0.08]	0.0
4	0.04 [0.01; 0.08]	0.0	0.02 [0.00; 0.03]	13.9	0.03 [0.00; 0.08]	1.3	0.06 [0.00; 0.16]	0.0
5					0.06 [0.01; 0.09]	0.0	0.05 [0.00; 0.13]	0.0
6					0.05 [0.04; 0.07]	0.7	0.02 [0.00; 0.04]	11.4
<i>Panel B: Europe</i>								
1	0.01 [0.00; 0.02]	12.0	0.02 [0.00; 0.04]	5.4	0.01 [0.00; 0.03]	6.5	0.02 [0.00; 0.04]	1.8
2	0.01 [0.00; 0.03]	8.6	0.01 [0.00; 0.03]	5.6	0.01 [0.00; 0.02]	20.5	0.02 [0.00; 0.05]	1.1
3	0.01 [0.00; 0.02]	14.7	0.01 [0.00; 0.03]	3.6	0.01 [0.00; 0.03]	5.5	0.01 [0.00; 0.04]	1.8
4	0.01 [0.00; 0.03]	9.3	0.01 [0.00; 0.01]	39.7	0.02 [0.00; 0.03]	5.2	0.02 [0.00; 0.04]	2.3
5					0.01 [0.00; 0.01]	38.0	0.01 [0.00; 0.02]	12.1
6					0.02 [0.00; 0.03]	5.8	0.01 [0.01; 0.02]	15.7
<i>Panel C: Japan</i>								
1	0.01 [0.00; 0.02]	33.0	0.01 [0.00; 0.02]	25.4	0.01 [0.00; 0.02]	35.5	0.01 [0.00; 0.02]	23.8
2	0.01 [0.00; 0.01]	50.7	0.02 [0.01; 0.04]	4.0	0.01 [0.00; 0.01]	45.5	0.01 [0.01; 0.04]	4.2
3	0.01 [0.00; 0.01]	40.1	0.01 [0.00; 0.02]	37.9	0.01 [0.00; 0.02]	23.8	0.02 [0.00; 0.04]	2.7
4	0.01 [0.00; 0.02]	27.6	0.01 [0.00; 0.02]	29.2	0.01 [0.00; 0.02]	29.9	0.01 [0.00; 0.02]	18.8
5					0.01 [0.00; 0.02]	40.5	0.00 [0.00; 0.01]	60.9
6					0.01 [0.00; 0.02]	24.8	0.01 [0.00; 0.03]	13.6
<i>Panel D: North America</i>								
1	0.01 [0.00; 0.03]	0.0	0.01 [0.00; 0.03]	0.1	0.02 [0.00; 0.04]	0.0	0.02 [0.00; 0.04]	0.0
2	0.01 [0.00; 0.02]	0.3	0.01 [0.00; 0.02]	0.7	0.02 [0.00; 0.06]	0.0	0.02 [0.00; 0.06]	7.0
3	0.01 [0.00; 0.02]	0.6	0.01 [0.00; 0.02]	0.0	0.01 [0.00; 0.02]	0.1	0.01 [0.00; 0.03]	0.2
4	0.02 [0.00; 0.03]	0.0	0.01 [0.00; 0.02]	0.0	0.01 [0.00; 0.02]	0.9	0.01 [0.00; 0.03]	0.1

(continued on next page)

Table A9 (continued)

Quartile/Sextile	Idiosyncratic risk quartile portfolios				Idiosyncratic risk sextile portfolios			
	Equally weighted		Mcap weighted		Equally weighted		Mcap weighted	
	Mean [min; max] of mean difference]	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value	Mean [min; max] of mean difference	T-test min p-value
5					0.01 [0.00; 0.03]	0.0	0.01 [0.00; 0.03]	0.0
6					0.03 [0.00; 0.05]	0.0	0.01 [0.00; 0.02]	0.0

Notes: The table provides the mean, minimum, and maximum of the absolute values of the mean differences in daily idiosyncratic risk per quartile and sextile portfolio and geographic region of the rating pairs in the period from January 2014 to December 2019. All values are presented in percent. Idiosyncratic risk per portfolio is computed by first running Eq. (3) on the daily portfolio returns of every month to estimate the betas in Eq. (3). Thereafter, the betas are used to compute the residual  $\epsilon_{i,t}$  for every observation. Idiosyncratic risk is then computed as the standard deviation of the residuals of every month. We compute the mean daily idiosyncratic risk per month for every quartile and sextile portfolio of the 36 rating pairs per geographic region. Thereafter, we compute the difference between mean idiosyncratic risk of the two matching quartile/sextile portfolios of a rating pair. The table shows the mean of the absolute values of these differences. The minimum and maximum values of these differences are shown in the squared brackets. In addition, we provide the minimum p-value of the t-tests per two matching quartile/sextile portfolios among the rating pairs. Portfolio returns are computed with equally weighted stock returns as well as with stock returns weighted by stocks' market capitalization. Example: When portfolio returns are computed with equally weighted returns, on average, the difference in mean daily idiosyncratic risk between the first quartile portfolio of a rating provider A and the first quartile portfolio of a provider B is 0.01 % for North American stocks, with a minimum value of 0.00 % and a maximum value of 0.03 %. The smallest p-value of the t-tests on the difference in idiosyncratic risk between the first quartile portfolio of a rating provider A and the first quartile portfolio of a provider B is 0.0 % for North American stocks. This shows that the idiosyncratic risk of the two most different portfolios is different from each other with a statistical significance at the 1 % level.

References

Adler, T., & Kritzman, M. (2008). The cost of socially responsible investing. *Journal of Portfolio Management*, 35(1), 52–56. <https://doi.org/10.3905/JPM.2008.35.1.52>

Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65(10), 4451–4469.

Alekseev, G., Giglio, S., Maingi, Q., Selgrad, J., & Stroebel, J. (2022). *A quantity-based approach to constructing climate risk hedge portfolios*. Working Paper.

Andersson, M., Bolton, P., & Samama, F. (2016). Hedging climate risk. *Financial Analysts Journal*, 72(3), 13–32.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 51, 259–299. <https://doi.org/10.1111/j.1540-6261.2006.00836.x>

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic risk and low returns: International and further US evidence. *Journal of Financial Economics*, 91, 1–23. <https://doi.org/10.1016/j.jfineco.2007.12.005>

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2023). Climate change concerns and the performance of green vs. brown stocks. *Management Science*, 69, 7151–7882.

Assaf, C., Monne, J., Harriet, L., & Meunier, L. (2024). ESG investing: Does one score fit all investors' preferences? *Journal of Cleaner Production*, 443, Article 141094. <https://doi.org/10.1016/j.jclepro.2024.141094>

Avramov, D., Cheng, S., Lioui, A., & Tarelli, A. (2022). Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, 145(2), 642–664.

Barnett, M. L., & Salomon, R. M. (2006). Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. *Strategic Management Journal*, 27(11), 1101–1122.

Bauer, R., Ruof, T., & Smeets, P. (2020). Get real! Individuals prefer more sustainable investments. *Review of Financial Studies*, 34(8), 3976–4043.

Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility? *Journal of Finance*, 73(6), 2471–2535.

Benson, K. L., & Humphrey, J. E. (2008). Socially responsible investment funds: Investor reaction to current and past returns. *Journal of Banking & Finance*, 32, 1850–1859. <https://doi.org/10.1016/j.jbankfin.2007.12.013>

Berg, F., Fabisik, K., & Sautner, Z. (2020). Rewriting history II: The (un) predictable past of ESG ratings. In *European Corporate Governance Institute-Finance Working Paper 708*, no. 2020 (pp. 10–2139).

Berg, F., Heeb, F., & Kölbl, J. (2024). The economic impact of ESG ratings. Working Paper [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4088545](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4088545).

Berg, F., Koelbel, J., Pavlova, A., & Rigobon, R. (2021). *ESG confusion and stock returns: Tackling the problem of noise*. Working Paper.

Berg, F., Koelbel, J., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344. <https://doi.org/10.1093/rof/rfac033>

Bessler, W., Taushanov, G., & Wolff, D. (2021). Optimal asset allocation strategies for international equity portfolios: A comparison of country versus industry optimization. *Journal of International Financial Markets Institutions and Money*, 72, Article 101343.

Bessler, W., & Wolff, D. (2024). Portfolio optimization with sector return prediction models. *Journal of Risk and Financial Management*, 17(254), 1–35.

Bialkowski, J., & Starks, L. T. (2016). *SRI funds: Investor demand, exogenous shocks, and ESG profiles*. Working Paper, University of Texas.

Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). Inside the ESG ratings: (dis)agreement and performance. *Corporate Social Responsibility and Environmental Management, special issue on Environmental, social, governance: Implications for businesses and effects for stakeholders*, 28(5), 1426–1445.

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517–549.

Bouslah, K., Kryzanowski, L., & M'Zali, B. (2018). Social performance and firm risk: Impact of the financial crisis. *Journal of Business Ethics*, 149, 643–669. <https://doi.org/10.1007/s10551-016-3017-x>

Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1–4), 21–58.

Brodback, D., Guenster, N., & Mezger, D. (2019). Altruism and egoism in investment decisions. *Review of Financial Economics*, 37(1), 118–148.

Cai, Y., Pan, C., & Statman, M. (2016). Why do countries matter so much in corporate social performance? *Journal of Corporate Finance*, 41, 591–609.

Chatterji, A. K., Durand, R., Levine, D. I., & Touboul, S. (2016). Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8), 1597–1614.

Clementino, E., & Perkins, R. (2021). How do companies respond to environmental, social and governance (ESG) ratings? Evidence from Italy. *Journal of Business Ethics*, 171, 379–397. <https://doi.org/10.1007/s10551-020-04441-4>

De Nard, G., Engle, R., & Kelly, B. (2024). Factor-mimicking portfolios for climate risk. *Financial Analysts Journal*, 80(3), 37–58.

DeMiguel, V., Garlappi, M., & Uppal, R. (2009). Optimal versus Naive Diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, 22(5), 1915–1953. <https://doi.org/10.1093/rfs/hhm075>

Diallo, B., Bagudu, A., & Zhang, Q. (2023). Fama–French three versus five, which model is better? A machine learning approach. *Journal of Forecasting*, 42(6), 1461–1475.

Dorantes, A. (2013). MVPOR: Stata module for collection, optimization and Backtest of financial portfolios. In *Statistical software components S457712*. Boston College Department of Economics.

Dorflleitner, G., Halbritter, G., & Nguyen, M. (2015). Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches. *Journal of Asset Management*, 16, 450–466.

Engle, R. (2024). *Termination risk and sustainability*. Working Paper.

Engle, R., Giglio, S., Kelly, B., Heebum, L., & Stroebel, J. (2020). Hedging climate change news. *Review of Financial Studies*, 33(3), 1184–1216.

Evans, J. L., & Archer, S. H. (1968). Diversification and the reduction of dispersion: An empirical analysis. *Journal of Finance*, 23(5), 761–767.

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

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

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance and Investment*, 5(4), 210–233.

Galema, R., & Gerritsen, D. (2022). ESG rating score revisions and stock returns. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4218969](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4218969).

Geczy, C., Stambaugh, R., & Levin, D. (2005). *Investing in socially responsible mutual funds*. Working Paper.

Gibson, R., Krueger, P., & Schmidt, P. (2021). ESG rating disagreement and stock returns. *Financial Analysts Journal*, 77(4), 104–127.

Giese, G., Lee, L.-E., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *The Journal of Portfolio Management*, 45, 69–83.

Giglio, S., Kelly, B., & Stroebel, J. (2021). Climate finance. *Annual Review of Financial Economics*, 13, 15–36.

- Giglio, S., Kuchler, T., Stroebel, J., & Zeng, X. (2023). Biodiversity risk. In *NBER Working Paper*.
- Giglio, S., Maggiori, M., Stroebel, J., Tan, Z., Utkus, S., & Xu, X. (2023). Four facts about ESG beliefs and investor portfolios. In *NBER Working Paper* 31114.
- Graham, J. (2022). Presidential address: Corporate finance and reality. *Journal of Finance*, 77, 1975–2049.
- Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *Journal of Finance*, 74(6), 2789–2837. <https://doi.org/10.1111/jofi.12841>
- Heeb, F., Kölbl, J., Paetzold, F., & Zeisberger, S. (2023). Do investors care about impact? *Review of Financial Studies*, 36(5), 1737–1787. <https://doi.org/10.1093/rfs/hhac066>
- Hervé, F., & Marsat, S. (2024). Acting for good, being good or feeling good? Exploring factors influencing individual investors' willingness to invest in green funds. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2024.105736> (forthcoming).
- Hoepner, A., Oikonomou, I., Sautner, Z., Starks, L., & Zhou, X. (2024). ESG shareholder engagement and downside risk. *Review of Finance*, 28(2), 483–510. <https://doi.org/10.1093/rof/rfad034>
- Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1), 15–36.
- Horn, M. (2023). The influence of ESG ratings on idiosyncratic stock risk: The unrated, the good, the bad, and the sinners. *Schmalenbach Journal of Business Research*, 75, 415–442.
- International Energy Agency. (2022). *World Energy Outlook 2022*.
- ISS ESG. (2023). ESG Corporate Rating: Methodology and Research Process. available at <https://www.issgovernance.com/file/products/iss-esg-corporate-rating-methodology.pdf>.
- Kotsantonis, S., & Serafeim, G. (2019). Four things no one will tell you about ESG data. *Journal of Applied Corporate Finance*, 31(2), 50–58.
- Krueger, P., Sautner, Z., & Starks, L. (2020). The importance of climate risks for institutional investors. *Review of Financial Studies*, 33(3), 1067–1111.
- Larcker, D., Tayan, B., & Watts, E. (2022). Seven Myths of ESG. *European Financial Management*, 28(4), 869–882.
- Latino, C., Pelizzon, L., & Rzeznik, A. (2021). The power of ESG ratings on stock markets. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3801703](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3801703).
- Lean, H., & Pizzutilo, F. (2020). Performances and risk of socially responsible investments across regions during crisis. *International Journal of Finance and Economics*, 26(3), 3556–3568.
- Levy, H. (1978). Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio. *American Economic Review*, 68(4), 643–658.
- Liang, H., & Renneboog, L. (2020). *Corporate social responsibility and sustainable finance: A review of the literature*. ECGI Finance Working Paper 701/2020.
- Lo, A. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.
- LSEG. (2022). Environmental, social and governance scores from LSEG. available at: [https://www.lseg.com/content/dam/data-analytics/en\\_us/documents/methodology/lseg-esg-scores-methodology.pdf](https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf).
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91.
- Merton, R. (1987). A simple model of capital market equilibrium with incomplete information. *Journal of Finance*, 42, 483–510.
- Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1), 31–42.
- MSCI. (2016). *MSCI ESG KLD STATS: 1991–2015 Data Sets*.
- Oehler, A., & Horn, M. (2021). *Contemporaneous ESG ratings and idiosyncratic stock risk: Empirical Evidence on Measures of Market Consensus and Dispersion*. Working Paper.
- Oehler, A., Horn, M., & Wendt, S. (2018). Why self-commitment is not enough: On a regulated minimum standard for ecologically and socially responsible financial products and services. In T. Walker, S. D. Kibsey, & R. Crichton (Eds.), *Designing a sustainable financial system: Development goals and socio-ecological responsibility* (pp. 405–421). Cham: Springer Nature.
- Pástor, L., Stambaugh, R., & Taylor, L. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics*, 142(2), 550–571.
- Pástor, L., Stambaugh, R., & Taylor, L. (2022). Dissecting green returns. *Journal of Financial Economics*, 146, 403–424.
- Pedersen, L., Fitzgibbons, S., & Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 142(2), 572–597. <https://doi.org/10.1016/j.jfineco.2020.11.001>
- Pedersen, L. H., Babu, A., & Ari, L. (2021). Enhanced portfolio optimization. *Financial Analysts Journal*, 77(2), 124–151.
- Raghuandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, 27, 822–863. <https://doi.org/10.1007/s11142-022-09693-1>
- Renneboog, L., Ter Horst, J., & Zhang, C. (2008). Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance*, 32, 1723–1742. <https://doi.org/10.1016/j.jbankfin.2007.12.039>
- Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *Journal of Finance*, 72(6), 2505–2550. <https://doi.org/10.1111/jofi.12547.7>
- Serafeim, G., & Yoon, A. (2022). Stock Price reactions to ESG news: The role of ESG ratings and disagreement. *Review of Accounting Studies*, 28, 1500–1530.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425–442.
- Sharpe, W. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 21(1), 49–58.
- Stambaugh, R. F., Yu, J., & Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *Journal of Finance*, 70(5), 1903–1948. <https://doi.org/10.1111/jofi.12286>
- Starks, L. (2021). Environmental, social, and governance issues and the financial analysts journal. *Financial Analysts Journal*, 77(4), 5–21.
- Starks, L. (2023). Presidential address: Sustainable finance and ESG issues – Value versus values. *Journal of Finance*, 78(4), 1837–1872.
- Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, 22(3), 353–363.
- Statman, M. (2000). Socially responsible mutual funds. *Financial Analysts Journal*, 30–39.
- Statman, M. (2004). The diversification puzzle. *Financial Analysts Journal*, 60(4), 44–53.
- Stroebel, J., & Wurgler, J. (2021). What do you think about climate finance? *Journal of Financial Economics*, 142(2), 487–498.
- Sustainalytics. (2019). *The ESG Risk Ratings*.
- Temple-West, P., & Schmitt, W. (2024). Investors pull cash from ESG funds as performance lags. *Financial Times* June 5 2024. Available at <https://on.ft.com/4bLQEkd>.
- Vigeo Eiris. (2020). *Corporate sustainability rating methodology*.
- Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34, 28–35. <https://doi.org/10.1093/biomet/34.1-2.28>
- Zerbib, O. (2022). A sustainable capital asset pricing model (S-CAPM): Evidence from green investing and sin stock exclusion. *Review of Finance*, 26(6), 1345–1388.