Are ESG ratings informative to forecast market risk ?

Christophe Boucher*, Wassim Le Lann†, Stéphane Matton‡, Sessi Tokpavi§

February, 2023

Abstract

Sustainable investing is growing fast and investors are increasingly integrating environmental, social, and governance (ESG) criteria. However, ESG ratings are derived using heterogeneous methodologies and can be quite divergent across providers, which suggests the need for a formal statistical procedure to evaluate their accuracy. This paper develops a backtesting procedure that evaluates how well these extra-financial metrics help in predicting a company's idiosyncratic risk. Technically, the inference is based on extending the conditional predictive ability test of Giacomini and White (2006) to a panel data setting. We apply our methodology to the forecasting of stock returns idiosyncratic volatility and compare two ESG rating systems from Sustainalytics and Asset4 across three investment universes (Europe, North America, and the Asia-Pacific region). The results show that the null hypothesis of no informational content in ESG ratings is strongly rejected in Europe, whereas results appear mixed in the other regions. Furthermore, the predictive accuracy gains are higher when considering the environmental dimension of ESG ratings. Importantly, applying the test only to firms over which there is a high degree of consensus between the ESG rating agencies leads to higher predictive accuracy gains for all three universes. Beyond providing insights into the accuracy of each of the ESG rating systems, this last result suggests that information gathered from several ESG rating providers should be cross-checked before ESG is integrated into investment processes.

JEL Codes: G10, G17, C12, C33.

Keywords: Backtesting, ESG ratings, ESG-related events, Idiosyncratic realised volatility, Test of equal predictive power, Panel data, Consensus ESG ratings.

*christophe.boucher@parisnanterre.fr, EconomiX-UPL, CNRS, University Paris Nanterre
†Corresponding author. w.lelann@outlook.fr, University of Orléans, LEO, Orléans, France.
‡stephane.matton@parisnanterre.fr, EconomiX-UPL, CNRS, University Paris Nanterre
§sessi.tokpavi@univ-orleans.fr, University of Orléans, LEO, Orléans, France.

1 Introduction

According to the Global Reporting Initiative (GRI), most of the ESG rating systems currently in use are designed to assess how effectively a company manages sustainability issues that have financial implications for its business. In other words, these systems evaluate a company's potential exposure to financial risks resulting from inadequate management of sustainability issues.[1] However, ESG ratings are derived using heterogeneous methodologies and can be quite divergent across rating agencies (Berg et al., 2020; Dimson et al., 2020), which raises concerns about their accuracy as a risk measure. Is there any informational content in the various existing ESG rating systems? Is this informational content related to what it is supposed to measure, which is the exposure of a company to sustainability-related risks ? There is clearly great interest in this issue as ESG is currently one of the most well-known acronyms in the financial world and beyond. Today, ESG ratings increasingly shape the investment decisions of investors. According to Bloomberg, ESG assets are on track to exceed $53 trillion by 2025, representing more than a third of projected total assets under management in North America, Europe, and Asia-Pacific capital markets.[2] This article aims to provide a statistical methodology to answer these questions by developing a backtesting procedure to assess the informational content of ESG ratings in forecasting a company's risk-related outcome. Our test evaluates the effectiveness of extra-financial metrics in predicting a company's risk exposure beyond the information conveyed by traditional financial variables.

The global craze for responsible investment has by now led to an abundant and rich literature that has tried, with mixed results, to evaluate how sustainable investment impacts market variables, and asset prices in particular. Some studies have found that ESG has a positive impact on asset prices (Mozaffar et al., 2016; Amel-Zadeh and Serafeim, 2018; Dyck et al., 2019; Hartzmark and Sussman, 2019), and Mozaffar et al. (2016) for instance present evidence that firms doing well on ESG issues outperform firms doing poorly on these issues. Amel-Zadeh and Serafeim (2018) reaffirm that ESG ratings have a material impact on asset prices and more specifically on the cost of capital, as investors expect higher return on equity for companies with strong ESG performance. Other contributions highlight that socially responsible investors can substantially reduce the cost of capital of responsible companies by tilting their portfolio allocation towards these firms (Gollier and Pouget, 2022; Zerbib, 2022). Dyck et al. (2019) also demonstrate that engagement by investors has a positive impact on ESG performance and ultimately on financial returns, especially in countries where ESG

[1] https://www.globalreporting.org/media/vyelrdub/gri-perspective-abc-of-esg-ratings-08.pdf

[2] https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by-2025-a-third-of-global-aum/

issues are important. A study of US mutual funds flows confirms that investors find value in sustainability as a positive predictor of future returns (Hartzmark and Sussman, 2019).

Arguing the other side though are some works (Riedl and Smeets, 2017; Pástor et al., 2021; Pedersen et al., 2020) based on the impact of investor preferences on the dynamics of asset prices (Fama and French, 2007), which report that ESG practices have either a negative or a positive impact on asset prices. Considering investor preferences for ESG, Riedl and Smeets (2017) notice that investors are willing to accept lower expected returns and higher management fees for holding companies with strong ESG performance. Pástor et al. (2021) model investor preferences for ESG in a mean-variance framework and show that in equilibrium, assets considered green generally have lower expected returns but provide greater utility and offer the ability to hedge against climate risk. They also introduce an ESG-factor that reacts to unexpected change in ESG, then conclude that green assets outperform when a positive shock hits this factor. Pedersen et al. (2020) extend the mean-variance-ESG framework by adding a third type of investor who is unaware of the ESG performance of firms. How the ESG ratings affect expected returns then depends on the wealth of this third investor.

Although this literature provides useful information on the link between extra-financial performance and asset price dynamics, it does not provide a formal methodology to assess whether the available rating systems are effective in measuring a company's exposure to financially material sustainability risks. This gap in the literature is all the more worrying as the correlations between the ratings of the various available providers are weak. Indeed, the divergence of ESG ratings has been widely documented (Chatterji et al., 2009; Semenova and Hassel, 2015; Chatterji et al., 2016; Berg et al., 2020; Dimson et al., 2020), and Berg et al. (2020) find, for instance, that correlations between the ESG ratings of providers are on average 61% for a set of five different ESG providers, whereas the credit ratings from the main agencies exhibit, on average, a correlation of 99%. They further explore the source of this divergence by splitting it into three components and looking at scope, or the selection of ESG categories to be measured; measurement, or how the ESG categories are assessed; and weight, or the importance given to each category. They observed that measurement explains more than 50% of the total divergence.[3] The divergence of ESG rating systems has important implications for sustainable investing. ESG ratings disagreement can lead to completely opposite opinions on one and the same company, dispersing ESG preferences of investors (Billio et al., 2019). It also makes it difficult to empirically assess the impact of ESG performance on stock returns (Berg et al., 2022) and can result in risk premiums for

[3]Unlike credit ratings, ESG ratings are most often created mainly from non-standardised information and are not regulated. Methodologies can be opaque and proprietary, leading to substantial rating divergence.

companies with high rating disagreement (Gibson Brandon et al., 2021).

Against this background, our paper introduces a statistical inferential procedure that allows to test the informational content of a given ESG rating system in forecasting a company's risk-related outcomes. The test is based on the idea that ESG ratings should have significant power in predicting the materialization of sustainability-related financial risks, as they are supposed to be informative about a company's exposure to such risks. Previous literature on the relationship between ESG ratings and firm-level risk outcomes has focused on two types of outcomes: ESG incidents and measures of financial risk. Several studies have found a link between ESG risks and idiosyncratic volatility (Jo and Na, 2012; Mishra and Modi, 2013; Bouslah et al., 2013; Sodjahnin et al., 2017; Hoepner et al., 2018; Albuquerque et al., 2019; Ilhan et al., 2019). For example, Mishra and Modi (2013) note that companies with lower leverage and high ESG ratings are better at capturing the benefits of ESG performance to reduce idiosyncratic risk.

Other studies, such as Champagne et al. (2019) and Serafeim and Yoon (2021), have examined the link between extra-financial performance or ESG ratings and the likelihood of adverse ESG events. Their analysis is based on the hypothesis that firms with strong extra-financial performance, such as good environmental externalities, employee relationships, and governance, are less likely to face ESG events such as environmental problems, employee claims, social conflicts, or boycotts and negative media campaigns. Champagne et al. (2019) use logistic regression to test whether a firm's extra-financial performance in a given year significantly helps anticipate ESG events in the following year. They find that an increase of one unit in a firm's rating reduces its probability of facing adverse events during the following year by 8%, controlling for financial performance. Similarly, Serafeim and Yoon (2021) investigate whether ESG ratings predict future ESG news and associated market reactions. Using a firm-day panel dataset, they find that the latest consensus ESG rating is associated with future ESG news, but the link weakens for firms over which there is large disagreement among raters.

Our contribution is related to these works, but differs in several aspects. First, these works do not provide a formal test to check the informational content of ESG ratings in forecasting firm-related risks, which is the purpose of this article. We test the informational content of ESG ratings using a dynamic forward-looking approach in an out-of-sample environment, which is consistent with the practice of institutions revising their ratings over time to incorporate new information on environmental, social, and governance practices. Second, our approach accounts for possible misspecification of the econometric model used to measure the relationship between ESG ratings and the outcome variable. This differs from the previous literature, where the correctness of the econometric model is critical to

establishing the existence of this link. Third, while previous studies identify significant correlations between ESG ratings and firm risks, they fail to quantify the improvement in model fit resulting from incorporating extra-financial information. Our method compares the predictive ability of nested models containing financial and extra-financial information, allowing for such quantification. Technically, our inferential procedure extends the conditional predictive ability test of Giacomini and White (2006) to a panel setting. We derive the Gaussian asymptotic distribution of the test statistic under weak assumptions. Monte Carlo simulations, performed under different types of model misspecification, demonstrate that our test has good small sample properties, with good size and increasing power as the number of firms and sample length increase.

On the empirical side, we apply our procedure to the forecasting of a company's market risk measured by the idiosyncratic volatility of its stock price. While in practice our test procedure can be applied to any target variable, we opt for a measure of market risk rather than an outcome measuring the materialization of ESG incidents for multiple reasons. First, measures of ESG incidents often rely on proprietary tools that can be divergent across providers. The rank correlations between ESG incidents from Sustainalytics and Asset4 for instance are weak at 43% for Europe, 43% for North-America and 34% for the Asia-Pacific region.[4] Second, as acknowledged by the GRI, most of existing ESG rating system seek to capture a company's financial exposure to poorly managed sustainability issues rather than its impact. This definition is in line with most of asset managers needs as the vast majority of them use ESG information for its materiality on investment performance (Amel-Zadeh and Serafeim, 2018). On the other side, while ESG incidents captured by negative news media coverage can have a substantial impact on stock prices, they are not always financially relevant for investors (Serafeim and Yoon, 2022). As a consequence, and consistent with what most of ESG rating agencies seek to capture, we opt for a direct measure of financial risk captured by the market risk of a company's stock price.

We conduct empirical applications to illustrate our methodology, using two leading ESG rating systems, Sustainalytics and Asset4, for Europe, North America and the Asia-Pacific region. Our results show that the null hypothesis of no informational content in ESG ratings is strongly rejected in Europe, whereas results appear mixed, and predictive accuracy gains are low in the other regions. Furthermore, we find that predictive accuracy gains are higher when assessing the environmental rating compared to other dimensions of ESG rating. Lastly, and importantly, we find that the predictive accuracy gains derived from ESG ratings increase with the level of consensus between rating agencies for all three universes. This

[4]These figures are computed over the period from January 2010 to October 2018.

final finding can be linked to that highlighted by Serafeim and Yoon (2021), who find that the market reaction to ESG news is moderated by the consensus rating. From a practical standpoint, our results provide crucial information for portfolio managers who integrates ESG rating to assess companies' risk profile, as we show that it is necessary to cross-check the information gathered from multiple ESG rating providers before integrating ESG into the management process.

The rest of the article is organised as follows. Section 2 describes our backtesting procedure for ESG ratings, focusing on the formulation of the null hypothesis, the construction of the test statistic and the analysis of its asymptotic distribution. Section 3 simulates the small sample properties of the test statistic under various settings, and empirical applications are considered in Section 4. The last section concludes the paper.

2 The backtesting procedure

This section gives a description of the backtesting procedure for evaluating statistically the informational content in ESG ratings. In the first part, we fix the notations and clearly define the null hypothesis of interest, while in the second part we provide the test statistic and its asymptotic distribution for inference.

2.1 Notations and the null hypothesis

To formulate the null hypothesis of our test, we consider an investment universe with n traded firms, and let $y_{i,t}$ denote the value at month t of a target variable that is intended to measure firm-specific risks. For instance, a socially motivated investor seeking to manage the environmental and social impact of their asset portfolio can use a variable $y_{i,t}$ that measures ESG incidents, such as the ones provided by well-known providers (Sustainalytics, Asset4, TrueValue Labs, etc.), to test whether ESG ratings help predict future corporate misconduct. On the other hand, investors who are interested in the materiality of ESG information on investment performance can use a target variable that measures a firm's specific exposure to financial risks, such as idiosyncratic volatility. Therefore, our framework is general, as it enables users to choose a target variable relevant to their investment objectives.

Let $x_{i,t}$ be a vector of length p in which the elements are innovations on p financial variables that measure the financial strength of firm i for the month t. Examples of such variables are dividend yield, sales over assets, debt over assets, or the quick ratio. They measure different facets of a firm's solvency including its size, returns, risk, liquidity, debt and leverage. Innovations can be obtained through autoregressive filtering on raw financial variables, or simply as deviations from the long-term average. Finally, the value of an ESG

rating is available for each firm i at month t and we denote it as $\omega_{i,t} \in \mathbb{R}$. This can be a global ESG rating measuring environmental, social and governance issues, or only one of these three components.

Now let $m^{(0)}_{i,t+\tau} = \mathbb{E}(y_{i,t+\tau} \,|x_{i,t})$ be the unknown expected value of $y_{i,t}$ for firm i at time $t+\tau$, conditional on its financial strength as measured by innovations $x_{i,t}$ in financial variables, with τ as a given forecast horizon. We can use a given predictive model, whether parametric, semi-parametric or non-parametric, to forecast $m^{(0)}_{i,t+\tau}$. The forecast we denote $\widehat{m}^{(0)}_{i,t+\tau}(\widehat{\beta}^{(0)}_{t,b_t})$ is based on the information set available at time t for all firms, so $\mathcal{F}^{(0)}_t = \{x_{i,s}, s = t-b_t+1, ..., t, i = 1, ..., n\}$, where b_t refers to the size of the estimation sample and $\widehat{\beta}^{(0)}_{t,b_t}$ collects all the estimated parameters. In a parametric model like a linear regression, $\widehat{\beta}^{(0)}_{t,b_t}$ is the vector of the estimates of the unknown parameters. Otherwise, it corresponds to whatever semi-parametric or non-parametric estimators are used to forecast $m^{(0)}_{i,t+\tau}$.

Let $m^{(1)}_{i,t+\tau} = \mathbb{E}(y_{i,t+\tau} \,|x_{i,t}, \omega_{i,t})$ be defined as $m^{(0)}_{i,t+\tau}$, but with the conditional set extended to $\omega_{i,t}$, so $\mathcal{F}^{(1)}_t = \{x_{i,s}, \omega_{i,s}, s = t-b_t+1, ..., t, i = 1, ..., n\}$. In other words, $m^{(1)}_{i,t+\tau}$ is the expected value of $y_{i,t}$ for firm i at time $t+\tau$, conditional on its financial states as given by $x_{i,t}$ and also on its ESG rating as given by $\omega_{i,t}$. We denote $\widehat{m}^{(1)}_{i,t+\tau}(\widehat{\beta}^{(1)}_{t,b_t})$ as the forecast value at time $t+\tau$.

Suppose that we produce T_0 out-of-sample forecasts of both the expected values $m^{(0)}_{i,t+\tau}$ and $m^{(1)}_{i,t+\tau}$ for each firm, so $\widehat{m}^{(0)}_{i,t+\tau}(\widehat{\beta}^{(0)}_{t,b_t})$ and $\widehat{m}^{(1)}_{i,t+\tau}(\widehat{\beta}^{(1)}_{t,b_t})$, $i = 1, ..., n$, $t+\tau = 1, ..., T_0$. With a loss function at hand that we denote $\mathcal{L}(.)$, we can evaluate the predictive performance of each model, generating two panels of losses as $\mathcal{L}^{(0)}_{i,t+\tau} \equiv \mathcal{L}^{(0)}_{i,t+\tau}(y_{i,t+\tau}, \widehat{m}^{(0)}_{i,t+\tau}(\widehat{\beta}^{(0)}_{t,b_t}))$ and $\mathcal{L}^{(1)}_{i,t+\tau} \equiv \mathcal{L}^{(1)}_{i,t+\tau}(y_{i,t+\tau}, \widehat{m}^{(1)}_{i,t+\tau}(\widehat{\beta}^{(1)}_{t,b_t}))$, where again $y_{i,t+\tau}$ is the value of $y_{i,t}$ for firm i at time $t+\tau$. From these panels, let $\Delta\mathcal{L}_{i,t+\tau} = \mathcal{L}^{(1)}_{i,t+\tau} - \mathcal{L}^{(0)}_{i,t+\tau}$ be the panel of loss differentials, $i = 1, ..., n$, $t = 1, ..., T_0$, and $\mu_i(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t})$ the expected value of the loss differentials for firm i.

Hence, our null hypothesis of overall equal predictive ability of the two forecasting models can be stated as :

$$\mathbb{H}_0 : \bar{\mu}(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t}) = 0, \tag{1}$$

with the alternative hypothesis being :

$$\mathbb{H}_1 : \bar{\mu}(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t}) < 0, \tag{2}$$

where $\bar{\mu}(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t})$ is defined as :

$$\bar{\mu}(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t}) = \frac{1}{n}\sum_{i=1}^{n} \mu_i(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t}). \tag{3}$$

This null hypothesis calls for several remarks. First, when it holds, it means that overall (for all i and t) including the ESG rating $\omega_{i,t}$ in the information set does not help for

forecasting $y_{i,t}$. In consequence, we should conclude that the ESG rating system investigated is void of information about $y_{i,t}$. Under the alternative hypothesis, considering the ESG rating in forecasting $y_{i,t}$, overall, gives real benefit across all firms and times.

Second, in contrast to the traditional framework for comparing predictive ability in Diebold and Mariano (1995) and West (1996), we can observe that the null hypothesis involves $\mu_i(\widehat{\beta}^{(0)}_{t,b_t}, \widehat{\beta}^{(1)}_{t,b_t})$, which depends on $\widehat{\beta}^{(0)}_{t,b_t}$ and $\widehat{\beta}^{(1)}_{t,b_t}$, which are the estimated values of the parameters instead of their population values. As discussed by Giacomini and White (2006) in a pure time series context, this helps preserve the finite sample behaviour of the estimators in the evaluation procedure, hence reflecting the effect of estimation uncertainty on the relative performance of the forecasts. This estimation uncertainty allows the comparison of nested forecasting models contrary to previous tests of predictive ability. However, they underline that adopting such a framework means remembering that the null hypothesis does not check the equal predictive ability of the competing models, but rather of the forecasting methods, where these methods include the models as well as the estimation procedures and the possible choices of estimation window.

This last remark means that some care is required in applying our test procedure to check for the validity of the null hypothesis in (1). First, the size of the estimation window should be kept fixed in the rolling window procedure ($b_t = b$) to ensure that parameter uncertainty does not vanish asymptotically. This naturally rules out an expanding window forecasting scheme, but allows for iterated or fixed schemes. Second, we should retain the same forecasting model and scheme and the same estimation window length to generate the forecasts $\widehat{m}^{(0)}_{i,t+\tau}(\widehat{\beta}^{(0)}_{t,b})$ and $\widehat{m}^{(1)}_{i,t+\tau}(\widehat{\beta}^{(1)}_{t,b})$. This is an important requirement, as it guarantees that the two forecasts diverge only by the set of information used, $\mathcal{F}^{(0)}_t$ or $\mathcal{F}^{(1)}_t$, the first of which excludes data on the ESG ratings for all firms.

2.2 Test statistic and asymptotic distribution

In this section, we provide the test statistic to check for the null hypothesis of a lack of informational content in an ESG rating system as expressed in (1). To do this we use the literature on comparing predictive ability in panel data settings (Davies and Lahiri, 1995; Timmermann and Zhu, 2019; Akgun et al., 2020). This literature considers extending the traditional predictive accuracy test for time series to a panel framework and it provides a test for overall equal predictive ability, meaning for all cross-sectional and time units as specified in (1), and also tests for joint equal predictive ability across cross-sectional units or time clusters.

Specifically, we draw on the framework of Akgun et al. (2020) who extend the test of

Diebold and Mariano (1995) to a panel data setting, considering the following test statistic based on the sample mean of loss differentials over time and units, so

$$\bar{\mu}_{n,T_0} = (nT_0)^{-1} \sum_{i=1}^{n} \sum_{t+\tau=1}^{T_0} \Delta\mathcal{L}_{i,t+\tau}, \tag{4}$$

and is given by

$$\mathcal{T}_{n,T_0} = \frac{\bar{\mu}_{n,T_0}}{\bar{\sigma}_{n,T_0}/\sqrt{nT_0}}, \tag{5}$$

where

$$\bar{\sigma}_{n,T_0} = n^{-1} \sum_{i=1}^{n} \sigma^2_{i,T_0}, \tag{6}$$

and $\sigma^2_{i,T_0} = var(\sqrt{T_0}\mu_i(\widehat{\beta}^{(0)}_{t,b}, \widehat{\beta}^{(1)}_{t,b}))$ is the long run variance of the ith time series of loss differentials.

As our null hypothesis is an extension to a panel setting of the unconditional predictive ability test of Giacomini and White (2006), rather than the one of Diebold and Mariano (1995), we need here assumptions that differ from those of Akgun et al. (2020), to establish the asymptotic distribution of the test statistic in (5).

Assumption 1 *For a given forecast horizon $\tau \geq 1$ and estimation window size $b < \infty$, suppose that (i) $\{(y_{i,t}, x'_{i,t}, \omega_{i,t})', t = 1, ..., T_0\}$ for a given i is mixing with ϕ of size $-r/(2r-2)$, $r \geq 2$, or α of size $-r/(r-2)$, $r > 2$; (ii) $\mathbb{E}\,|\Delta\mathcal{L}_{i,t+\tau}|^{2r} < \infty$ for all t and a given i; (iii) $\sigma^2_{i,T_0} = var(\sqrt{T_0}\mu_i(\widehat{\beta}^{(0)}_{t,b}, \widehat{\beta}^{(1)}_{t,b})) > 0$ for all T_0 sufficiently large and a given i.*

Assumption 2 *$\bar{\mu}_{i,T_0} = T_0^{-1} \sum_{t+\tau=1}^{T_0} \Delta\mathcal{L}_{i,t+\tau}$, $i = 1, ..., N$ are independent, and*

$$\mathbb{E}(|\bar{\mu}_{i,T_0} = T_0^{-1} \sum_{t+\tau=1}^{T_0} \Delta\mathcal{L}_{i,t+\tau}|)^{2+\delta} < C < \infty, \tag{7}$$

for some $\delta > 0$ for all i. $\bar{\sigma}^2_{n,T_0} = n^{-1} \sum_{i=1}^{n} \sigma^2_{i,T_0} > \delta' > 0$ for all n sufficiently large.

Assumption 1 includes regularity conditions for the validity of Theorem 4 in Giacomini and White (2006). These conditions ensure that the test statistic for the unconditional predictive ability applied to a fixed cross-sectional unit converges to a standard Gaussian distribution, with

$$\mathcal{T}_i = \frac{\bar{\mu}_{i,T_0}}{\sigma_{i,T_0}/\sqrt{T_0}} \xrightarrow[T_0 \to \infty]{\mathcal{D}} N(0,1). \tag{8}$$

Assumption 2 assumes the independence between the n random variables $\bar{\mu}_{i,T_0}$, $i, ..., n$, meaning the average values over time of the loss differentials for each firm. This assumption allows the Central Limit Theory (CLT) applied to independent and heterogeneous random variables (White, 2001, Theorem 5.10) to hold. Note that this assumption is not a strong

one within our framework, as opposed to macroeconomic forecasting. Indeed, our focus is on target variables that are related to firm-specific risk, which is by its nature a specific measure for each firm and hence primarily driven by firm characteristics rather than common factors. The following proposition provides the asymptotic distribution of the test statistic in (5).

Proposition 1 *Under the null hypothesis of a lack of informational content in ESG ratings as stated in (1), and if Assumptions 1-2 hold, we have that*

$$\mathcal{T}_{n,T_0} = \frac{\bar{\mu}_{n,T_0}}{\bar{\sigma}_{n,T_0}/\sqrt{nT_0}} \xrightarrow[T_0,n\to\infty]{\mathcal{D}} N(0,1). \tag{9}$$

Thus we reject the null hypothesis when $\mathcal{T}_{n,T_0} < z_\eta$ with z_η the quantile of order η of the standard Gaussian distribution, and η the nominal significance level. The proof of Proposition 1 is straightforward following Akgun et al. (2020), as we may note that under $\mathbb{H}_0$,

$$\sqrt{nT_0}\bar{\mu}_{n,T_0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sqrt{T_0}\bar{\mu}_{i,T_0}, \tag{10}$$

with $\bar{\mu}_{i,T_0}$ as defined in Assumption 2. For a fixed i, if Assumption 1 holds, $\sqrt{T_0}\bar{\mu}_{i,T_0} \xrightarrow[T_0\to\infty]{\mathcal{D}} \psi_i$, with $\psi_i \sim N(0,\sigma^2_{i,T_0})$, and $\sigma^2_{i,T_0} = var(\sqrt{T_0}\mu_i(\widehat{\beta}^{(0)}_{t,b},\widehat{\beta}^{(1)}_{t,b}))$. See Theorem 4 in Giacomini and White (2006). Hence the rest of the proof proceeds by noting that under Assumption 2, the CLT for heterogeneous but independent variables (White, 2001, Theorem 5.10) holds and $(1/\sqrt{n})\sum_{i=1}^{n}\psi_i \xrightarrow[T_0,n\to\infty]{\mathcal{D}} N(0,\bar{\sigma}^2_{n,T_0})$, where again $\bar{\sigma}^2_{n,T_0} = n^{-1}\sum_{i=1}^{n}\sigma^2_{i,T_0}$.

Note that to compute our test statistic $\mathcal{T}_{n,T_0}$, we need a consistent estimate $\widehat{\bar{\sigma}}^2_{n,T_0}$ of $\bar{\sigma}^2_{n,T_0}$. Under the assumption of cross-sectional independence of loss differentials, it follows that $\widehat{\bar{\sigma}}^2_{n,T_0} = n^{-1}\sum_{i=1}^{n}\widehat{\sigma}^2_{i,T_0}$, where $\widehat{\sigma}^2_{i,T_0}$ is a suitable HAC estimator of the long-run variance σ^2_{i,T_0} of the ith time series of loss differentials, with

$$\widehat{\sigma}^2_{i,T_0} = T_0^{-1}\sum_{t+\tau=1}^{T_0}\Delta\mathcal{L}^2_{i,t+\tau} + 2[T_0^{-1}\sum_{j=1}^{p_{T_0}} w_{T_0,j}\times\sum_{t+\tau=1+j}^{T_0}\Delta\mathcal{L}_{i,t+\tau}\Delta\mathcal{L}_{i,t+\tau-j}], \tag{11}$$

and $\{p_{T_0}\}$ is a sequence of integers such that $p_{T_0}\to\infty$ as $T_0\to\infty$, $p_{T_0} = o(T_0)$, and $\{w_{T_0,j} : T_0 = 1,2,...; j = 1,...,p_{T_0}\}$ is a triangular array such that $|w_{T_0,j}| < \infty$, $T_0 = 1,2,...$; $j = 1,...,p_{T_0}$, $w_{T_0,j}\to 1$ as $T_0\to\infty$ for each $j = 1,...,p_{T_0}$ (Andrews, 1991).

3 Small sample properties

In this section we use a realistic simulation framework to analyse the small sample properties of the test. We begin by describing the simulation setup and then provide results for the sizes and the powers of the test under different forms of misspecification for the forecasting method retained.

3.1 Simulation setup

Our simulation setup proceeds by first simulating a vector $x_{i,t}$ of length $p = 10$ of innovations in financial variables that measure the financial strength of firm i at time t, with $t = 1, ..., T$ and $T \in \{120, 180, 240\}$ as the sample size corresponding to 12, 15 and 20 years of monthly data. To have a realistic setup, these p variables are generated from a multivariate Gaussian distribution with mean vector $\bar{x}$ and covariance matrix Ω calibrated using real data (see Appendix A for details about the calibration). With the vector $x_{i,t}$ of length p ready at hand, we generate the logarithmic value of the target variable $y_{i,t}$ for firm i, as[5] :

$$\log(y_{i,t+1}) = c_i^\star + x'_{i,t}\beta_i^\star + \gamma\omega_{i,t} + u_{i,t+1}, \tag{12}$$

with $u_{i,t+1}$ following a standard Gaussian distribution, $c_i^\star$ as the constant term and $\beta_i^\star$ as a vector of parameters of length p. Note that we allow for heterogeneity across firms with specific values for the parameters for each firm. The values of $c_i^\star$ are generated as follows :

$$c_i^\star = c^\star + U(-|\frac{c^\star}{10}|; |\frac{c^\star}{10}|), \tag{13}$$

with $U(a; b)$ as a uniform random variable over the set $[a, b]$. The same perturbation principle is used to generate each component of the vector $\beta_i^\star$, with :

$$\beta_{i,j}^\star = \beta_j^\star + U(-|\frac{\beta_j^\star}{10}|; |\frac{\beta_j^\star}{10}|), \tag{14}$$

$j = 1, ..., p = 10$. The reference values $c^\star$ and $\beta^\star$ of the parameters are calibrated using real data (see Appendix A for details).

In equation (12), $\omega_{i,t}$ is the ESG rating, which for firm i and at each date t is generated from a uniform distribution over the set $[0, 1]$, and $\gamma \in \mathbb{R}_-$ is a parameter. Note that our null hypothesis holds for $\gamma = 0$, since the ESG rating does not have any predictive content for $y_{i,t}$. With γ diverging from zero, the null hypothesis does not hold, because high lagged values of the ESG rating decrease the values of $y_{i,t}$.

Based on our design and for each Monte Carlo replication, with n and T fixed, the above simulation design is run for the n firms, with $n \in \{100, 250, 500\}$. This leads to a pure heterogeneous panel for $y_{i,t}$, the $p = 10$ innovations in financial variables $x_{i,t}$, and the ESG rating $\omega_{i,t}$, with $i = 1, ..., n$ and $t = 1, ..., T$.

3.2 Sizes and powers under a medium level of misspecification

For each Monte Carlo replication, we use the generated variables $y_{i,t}$, $x_{i,t}$ and $\omega_{i,t}$, $i = 1, ..., n$, $t = 1, ..., T$ and a fixed forecasting method to generate the forecast of $m_{i,t+1}^{(0)} = \mathbb{E}(y_{i,t+1} \,|x_{i,t})$

[5] We use the logarithm, as most of possible candidate variables for $y_{i,t}$ are positive, including ESG incident variables.

and $m^{(1)}_{i,t+1} = \mathbb{E}(y_{i,t+1}\,|x_{i,t}, \omega_{i,t})$, so $\widehat{m}^{(0)}_{i,t+1}(\widehat{\beta}^{(0)}_{t,b})$ and $\widehat{m}^{(1)}_{i,t+1}(\widehat{\beta}^{(1)}_{t,b})$ with b the estimation sample that we set to $b = [0.75T]$, and $[a]$ the integer part of a. This means that we use the first 75% of the T observations for each firm as the estimation sample, and generate T_0 forecasts corresponding to the last 25% of the observations, meaning $T_0 = [0.25T]$ and $T = T_0 + b$.

The forecasts for both models are obtained using pooled OLS regression models. This means that both forecasting models are misspecified, because the true panel structure of the data is heterogeneous across units. Besides, there is another form of misspecification that arises because the true data generating process uses a linear form for the *logarithm* of $y_{i,t}$ (see Eq. 12), while the pooled OLS regression models are fitted for the *raw* values of the same variable. Our goal is to evaluate how robust our inferential procedure is to these two levels of misspecification, which we call medium in comparison to another more severe form of misspecification that we will consider next. It may be recalled that the asymptotic behaviour of our test statistic under the null hypothesis suggests that with $\gamma \in \mathbb{R}_-$ in (12) diverging from zero, the null hypothesis is more likely to be rejected for $T_0, n \to \infty$, or equivalently, $T, n \to \infty$.

Figure 1 displays the rejection frequencies of the null hypothesis with respect to the parameter γ for a given couple (n, T), with the nominal significance level set to 5%. The rejection frequencies are computed over $1,000$ simulations. Overall the test exhibits very good small sample properties, and we observe that the rejection frequencies for all couples (n, T) are close to 5% for $\gamma = 0$ and increase monotonically as γ diverges from 0.

We also observe that for a fixed n and $\gamma < 0$ the powers increase with T. Indeed, for $n = 100$ and $\gamma = -0.25$, the rejection frequencies for T of 120, 180 and 240 are 39.10%, 53.30% and 61.00% respectively. The same behaviour is observed for a fixed T and $\gamma < 0$ with the powers increasing with n. For instance with $T = 120$ and $\gamma = -0.25$, the rejection frequencies for $n = 100$, 250 and 500 are respectively 39.10%, 71.50% and 91.30%. Hence our inferential procedure exhibits very good small sample properties. Figure B.1 in Appendix B displays the rejection frequencies for the same simulation setup using the absolute error loss function. We can observe similar small sample properties, offering proof that our test is robust to the loss function.

3.3 Sizes and powers under a high level of misspecification

We now consider a configuration that will help us evaluate the properties of the test with respect to the choice of financial variables. In the last subsection we assumed that the user of the test includes in the forecast models all the $p = 10$ innovations in the financial variables that enter the specification of the true model, but we make here the assumption

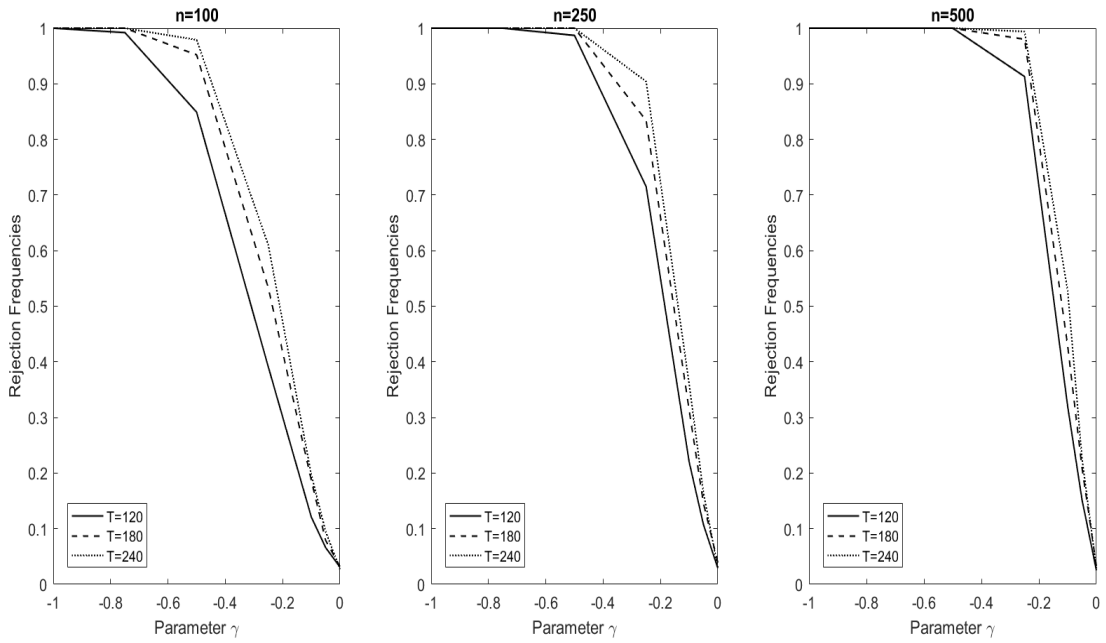


Figure 1: Rejection frequencies under a medium level of misspecification with the squared error loss function

that only some of these variables are retained. In each Monte Carlo replication, the following two pooled OLS models are estimated to compute out-of-sample forecasts $\widehat{m}_{i,t+1}^{(0)}(\widehat{\beta}_{t,b}^{(0)})$ and $\widehat{m}_{i,t+1}^{(1)}(\widehat{\beta}_{t,b}^{(1)})$ of $m_{i,t+1}^{(0)} = \mathbb{E}(y_{i,t+1}\,|x_{i,t})$ and $m_{i,t+1}^{(1)} = \mathbb{E}(y_{i,t+1}\,|x_{i,t}, \omega_{i,t})$:

$$y_{i,t+1} = c + \widetilde{x}_{i,t}'\beta + v_{i,t+1}^{(0)}, \tag{15}$$

$$y_{i,t+1} = c + \widetilde{x}_{i,t}'\beta + \omega_{i,t}\gamma + v_{i,t+1}^{(1)}, \tag{16}$$

with $v_{i,t+1}^{(0)}$ and $v_{i,t+1}^{(1)}$ as the error terms and $\widetilde{x}_{i,t}$ as a vector with $p/2$ randomly chosen financial variables from the $p = 10$ relevant ones as its elements, and $\widehat{\beta}_{t,b}^{(0)} = (\widehat{c}, \widehat{\beta}')'$, $\widehat{\beta}_{t,b}^{(1)} = (\widehat{c}, \widehat{\beta}', \widehat{\gamma})'$. Assessing the small sample properties of the test with this additional form of misspecification is of great interest because such misspecification could probably arise in empirical applications where users are very likely to be wrong in their choice of the financial variables that matter.

Figure 2 displays the rejection frequencies over $1,000$ simulations. We observe that the proposed test is robust to this form of misspecification. Indeed, the rejection frequencies are similar to those displayed in Figure 1, suggesting that making a mistake in the choice of financial variables is not harmful. Results available from the authors upon request show that the robustness holds even when the misspecification is more pronounced as only a quarter of the financial variables of interest are chosen. The robustness to the choice of the loss function can be seen in Figure B.2 in Appendix B.

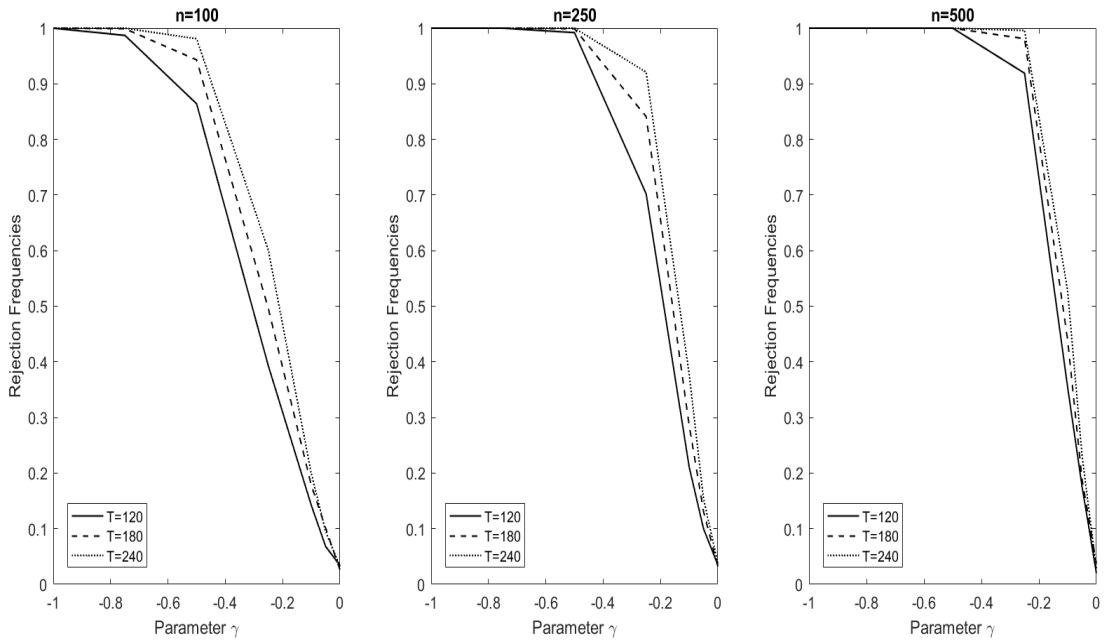


Figure 2: Rejection Frequencies under a high level of misspecification with the squared error loss function

4 Empirical applications

This section illustrates our backtesting procedure using real datasets. We apply our methodology to two popular providers of ESG ratings, Sustainalytics and Asset4, over three universes from North America, Europe and the Asia-Pacific region. We first describe our datasets and the related variables, and then conduct inferences to evaluate the informational content of each of the rating systems.

4.1 Description of the datasets and variables

The dataset for each of the three universes contains information for n firms at a monthly frequency over a period ranging from January 2010 to October 2018, giving a total of $T = 106$ months. Note that we restrict our investigations to this period, as Sustainalytics has made a major change in the methodology for constructing its ratings in December 2018, with an inconsistency in the chaining of the ratings before and after this date. Precisely, before (after) this date, the ratings are performance (risk) measures, with higher (lower) ratings corresponding to best practices for environmental, social and governance issues. Obviously, one solution would be to transform the risk-ratings into performance-ratings, but such a transformation would be arbitrary, and would not guarantee consistency in the scales of values. The North America, Europe and Asia-Pacific datasets gather information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. This deep panel structure ensures a

high power for our backtesting methodology (see Monte Carlo simulations), with a total of 34,556, 25,228 and 23,002 pooled observations for the North America, Europe and Asia-Pacific universes.

4.1.1 Information on ESG data

Table 1 displays pooled descriptive statistics of the ESG ratings for the two providers over the three universes. We may note that for both providers, higher values of the ESG ratings indicate higher ESG performance.

Table 1: Pooled descriptive statistics of the ESG ratings

	Min.	Max.	Mean	Median	Std.
			Europe		
Sustainalytics	36.0000	89.6900	66.5310	67.3000	9.6449
Asset4	5.4700	94.1500	64.4389	66.1300	15.7645
			North America		
Sustainalytics	33.0000	88.0000	59.0831	59.0000	8.6864
Asset4	2.4700	94.7700	54.4304	56.5200	18.8691
			Asia-Pacific		
Sustainalytics	32.0000	90.0900	58.5848	59.0000	8.3848
Asset4	2.3500	90.2700	53.3590	56.1900	18.2707

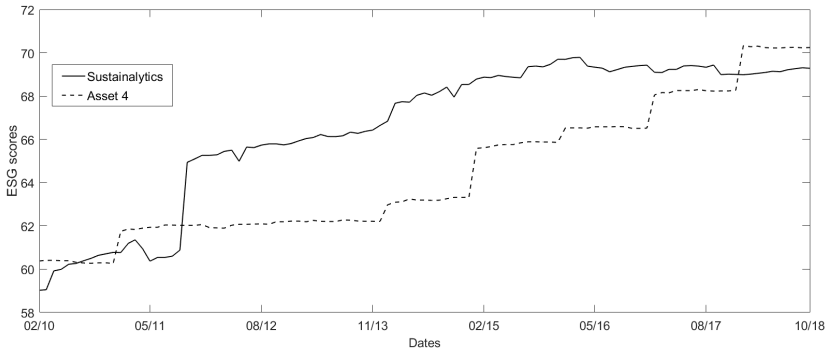
Notes: The table displays pooled descriptive statistics of the ESG ratings for the two providers (Sustainalytics and Asset4) over the three universes. The datasets contain monthly observations over the period from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. Min. refers to minimum, Max. to maximum, and std. to standard deviations.

The average values of the ESG ratings for the Europe universe are 66.53 for Sustainalytics and 64.43 for Asset4. This means the central statistics are similar for both providers, as is confirmed by the values of the median of 67.30 for Sustainalytics and 66.13 for Asset4 for the Europe universe. This stylised fact holds for the other two universes. However, the Asset4 ESG ratings have more variability across time and firms as given by the values of the standard deviations and ranges. The standard deviations of the Asset4 ESG ratings for instance are approximately twice as high as those for Sustainalytics.

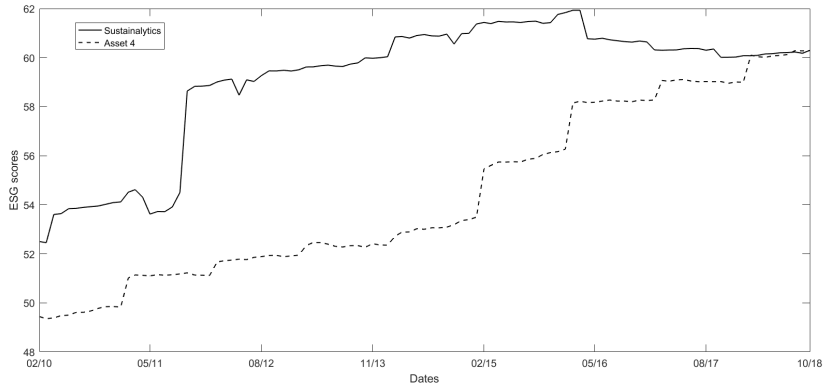
Figure 3 displays the evolution over time of the cross-sectional averages of the ESG ratings for the two providers in the three universes. We observe growth over time in the cross-sectional averages, which suggests a tendency towards upward revisions of the ESG ratings for firms. Assuming that ESG ratings accurately reflect ESG performance, this

Figure 3: Dynamics of the cross-sectional means of the ESG ratings

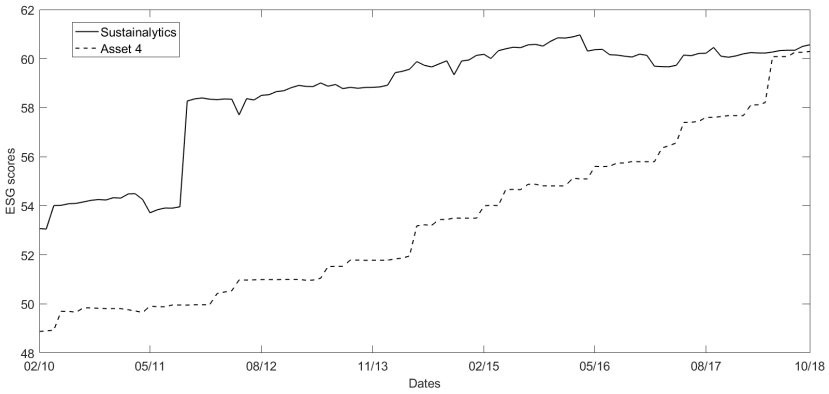
(a) Europe



(b) North America



(c) Asia-Pacific

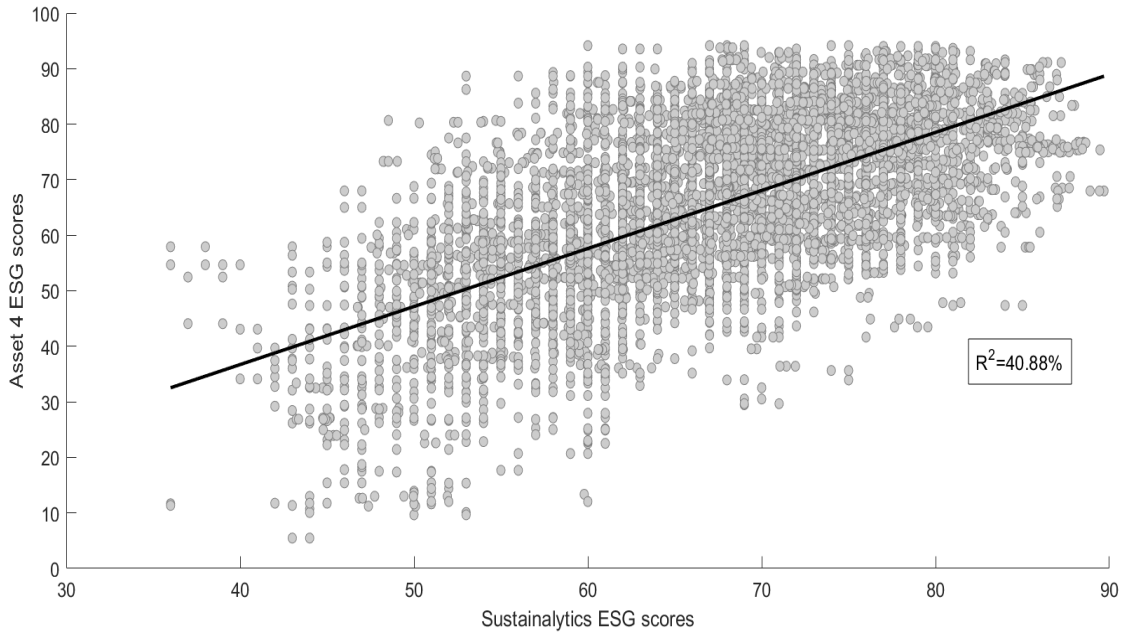


Source: The figure displays the evolution over time of the cross-sectional means of the ESG ratings for the two providers considered (Sustainalytics and Asset4). The dataset contains monthly observations for $n = 238$ firms from January 2010 to October 2018, giving a total of 106 months.

shows an overall improvement trend over time in the corporate behavior of firms across the three universes regarding environmental, social, and governance best practices.

To evaluate the link between the two rating systems, Figure 4 displays the scatter plot of the pooled ESG ratings from the two providers for the Europe universe. The figure also displays the fitted least square regression line, along with the adjusted R-squared, which is equal to 40.88%. Hence the link across firms and time between the two ESG ratings is weak, though it is positive. As already underlined, this has been highlighted many times in the literature and constitutes the main motivation of our paper, which proposes, in a context of limited convergence, a formal backtesting procedure for evaluating the informational content of ESG rating systems. The phenomenon is not only European and is also highlighted for the other two universes as shown by Figures B.3 and B.4 in Appendix B. The trend is of the same order for the North America universe with an R-squared of 46.46%, but we observe a more pronounced divergence in the Asia-Pacific universe with an R-squared of only 32.65%.

Figure 4: Relation between the Sustainalytics and Asset4 ESG ratings: Europe



Source: The figure displays the scatter plot that shows the graphical relation between the ESG ratings for the two providers considered (Sustainalytics and Asset4). The dataset contains monthly observations for $n = 238$ firms from January 2010 to October 2018, giving a total of 106 months.

4.1.2 Information on the target variable

In this sub-section, we provide information on the target variable. We consider the idiosyncratic volatility of stock returns as our dependent variable of interest. This variable measures market risk at the firm level that is not captured by traditional risk factors. ESG

ratings could significantly help predict this target variable as stock markets can react to the arrival of firm-specific ESG events (Serafeim and Yoon, 2021) or global news corresponding to innovations in an ESG factor (Pástor et al., 2021; Ardia et al., 2022).

Another choice could be a variable or score measuring ESG incidents from leading providers. However, they seem divergent across providers, as the rank correlations between ESG incidents from Sustainalytics and Asset4 for instance are weak at 43% for Europe, 43% for North America and 34% for the Asia-Pacific region. Moreover, in this paper, we adopt the perspective of an investor using ESG information for its materiality on investment performance because this is the primary reason why investors use ESG information and many rating agencies adopt this perspective (Amel-Zadeh and Serafeim, 2018). Since ESG news is not always financially relevant for investors (Serafeim and Yoon, 2022), using a direct measure of financial risks, such as idiosyncratic volatility, seems more appropriate in our context.

To compute idiosyncratic volatility for each firm i, we collect daily stock returns $r_{i,s}$ over our period of investigation from January 2010 to October 2018, with a total of $2,304$ observations. For each universe, we also collect the daily returns $r_{m,s}$ of the MSCI stock index over the same period, using MSCI Europe, MSCI USA and MSCI Pacific for the Europe, North America and Asia-Pacific universes. Residual returns are thus extracted assuming that the Capital Asset Pricing Model (CAPM) holds, with:

$$r_{i,s} = \alpha_i + \beta_i r_{m,s} + \epsilon_{i,s}, \tag{17}$$

where α_i is the alpha of the stock, β_i is the beta or exposure of the stock to the market, and $\epsilon_{i,s}$ is the innovation or residual return for stock i at day s. With the daily residual returns, we compute monthly idiosyncratic realized volatility as follows:

$$IRV_{i,t} = \sum_{s_k=1}^{v_t} \widehat{\epsilon}_{i,s_k}^2, \tag{18}$$

with t the index of the month, v_t the number of daily observations in month t, and $\widehat{\epsilon}_{i,s_k}$ the s_k^{th} fitted residual returns within month t. For each firm i in a given universe, we obtain a time series of monthly idiosyncratic realized volatility of length 106, which thus matches the monthly frequency and the length of the ESG data analysed in the previous sub-section. The backtesting procedure is then applied using the logarithmic transform of the idiosyncratic realized volatility as the target variable.

Remark 1 *The CAPM model in* (17) *is likely to be misspecified. In this case, our target variable* $y_{i,t} \equiv \log(IRV_{i,t})$ *would be correlated across firms. However, recall that Proposition 1 does not require cross-sectional independence between* $y_{i,t}$*, but rather between loss*

differentials averaged over time. Besides, we further use a multi-factorial model to check the sensitivity of our results to the choice of the factor model.

Table 2: Pooled descriptive statistics of idiosyncratic realised volatility

	Min (%)	Max (%)	Mean (%)	Median (%)	Std (%)
Europe	0.0133	24.6450	0.4634	0.2927	0.7064
North America	0.0099	50.6339	0.4970	0.2632	0.9217
Asia-Pacific	0.0236	33.3244	0.5416	0.3697	0.7303

Notes: The table displays pooled descriptive statistics of monthly idiosyncratic realised volatilities for the three universes. Idiosyncratic realised volatilities are computed from residual asset returns from the CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. Min. refers to minimum, Max. to maximum, and std. to standard deviations.

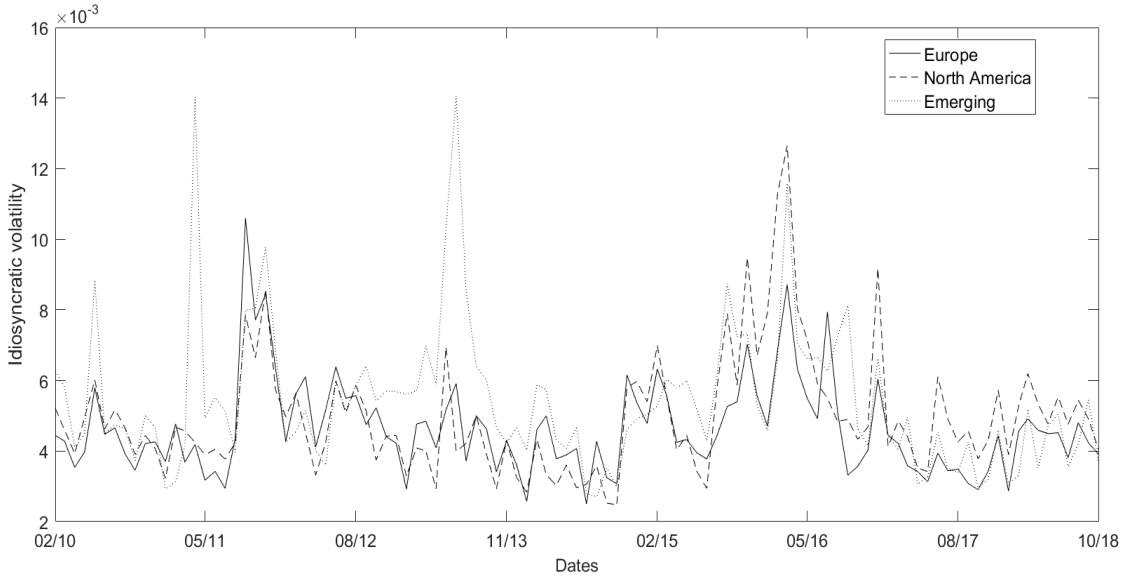
Table 2 displays the pooled descriptive statistics of monthly idiosyncratic volatilities for the three universes. The Asia-Pacific universe appears as the one where firms have on average the highest levels of idiosyncratic volatility. In terms of dispersion, the North America universe has more variability in the measure of the volatility of residual returns, as given by the values for the standard deviation and the range.

To get an overhead view of the monthly series of idiosyncratic realised volatilities, Figure 5 displays the evolution over time of the cross-sectional means of monthly idiosyncratic realised volatilities. We observe the typical dynamics, with volatility clusters that nevertheless seem less pronounced because we are dealing with idiosyncratic volatility, and not total volatility which includes the systematic part.

It may be recalled that our backtesting procedure is designed to test the informational content of the ESG ratings by checking whether they have predictive power for future market risk, as measured here by increased idiosyncratic volatility of stock returns. Hence, the relationship that the test aims to validate is that high ESG ratings lead to low idiosyncratic volatilities and low ratings lead to high volatilities.

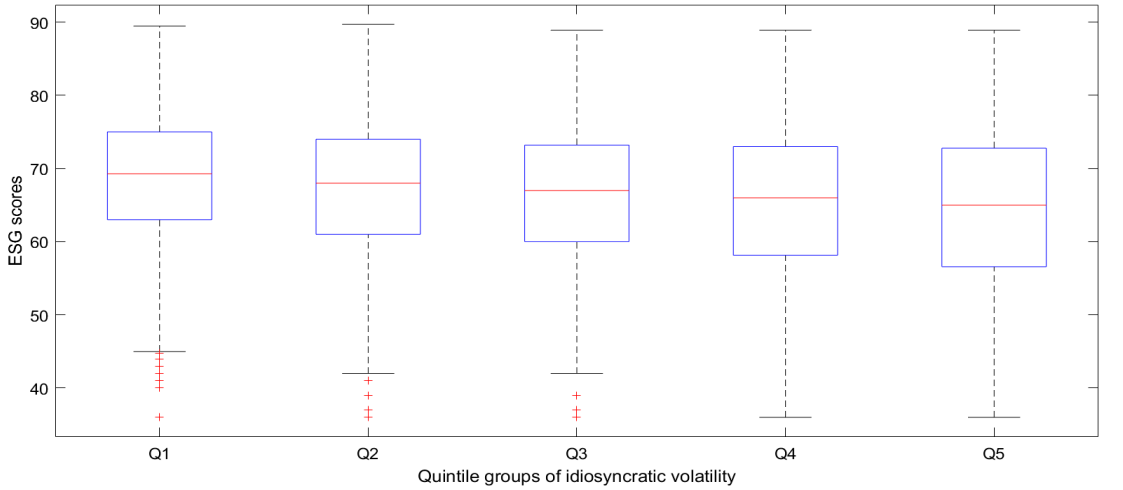
So before we apply the backtesting procedure formally, Figures 6 and 7 try to illustrate whether there is such a relationship in the Europe universe. These figures report the distribution of the lagged values of the ESG ratings (Figure 6 for Sustainalytics and Figure 7 for Asset4) by idiosyncratic volatility quintiles. Overall we observe that a negative relation arises, with high values of lagged ESG ratings associated with low idiosyncratic volatilities, while the median values of the lagged ESG ratings decrease with the order of the quintiles. Robustness across the universes is confirmed in Appendix B, with Figures B.5 and B.6 for the North America universe, and B.7 and B.8 for the Asia-Pacific universe.

Figure 5: Dynamics of the cross-sectional means of idiosyncratic realised volatility



Source: The figure displays the evolution over time of the cross-sectional means of monthly idiosyncratic realised volatilities. Idiosyncratic realised volatilities are computed from residual stock returns from the CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets contain information on respectively $n = 326$, $n = 238$ and $n = 217$ firms.

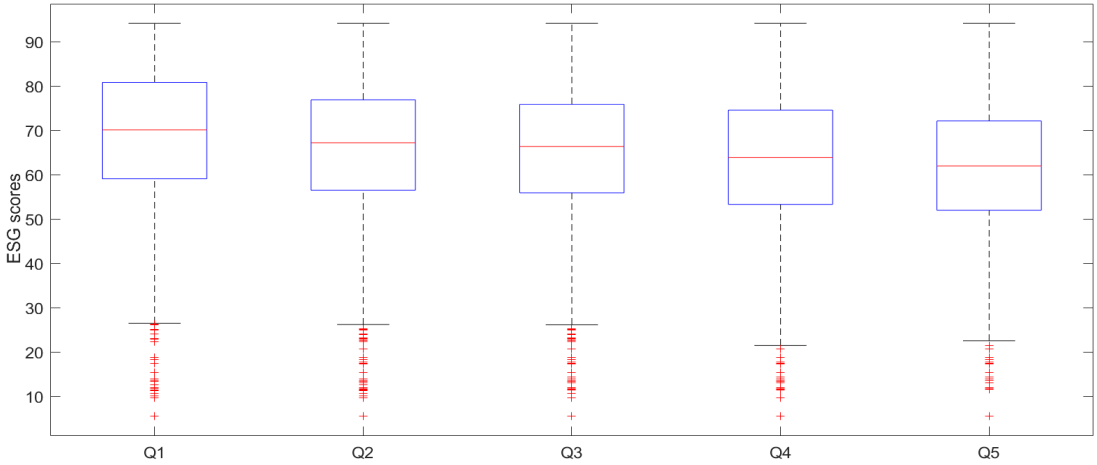
Figure 6: ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (Europe)



Source: For the Europe universe, the figure displays the means of the Sustainalytics ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 238$ firms from January 2010 to October 2018, giving a total of 106 months.

To control for potential confounding factors of the link between ESG ratings and idiosyncratic volatility, retain $p = 10$ financial variables for which the monthly observations are available for all firms over the three universes and the timespan considered. These vari-

Figure 7: ESG ratings by idiosyncratic volatility quintiles: Asset4 (Europe)



Source: For the Europe universe, the figure displays the means of Asset4 ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 238$ firms from January 2010 to October 2018, giving a total of 106 months.

ables are tax burden, interest burden, operating margin, asset turnover, leverage, current ratio, net debt to earnings before interest, taxes, depreciation, and amortisation (EBITDA), capital expenditure (Capex) to depreciation, current assets, and current liabilities (see table 3 for a complete description of these variables). Innovations are extracted for each of these financial variables and for each firm by centering the raw values on the time average.

Table 3: Description of financial variables

Variables	Ratios	Description
Tax Burden	Net Income/Pretax Income	Profits retained after taxes
Interest Burden	Pretax Income/EBIT	Profits retained after interest paid
Operating Margin	EBIT/Revenue	Return on sales
Asset Turnover	Revenue/Total Assets	Revenue generated by own resources
Leverage	Total Assets/Total Equity	Measure of financial leverage
Current Ratio	Current Assets/Current Liab.	Measure of short-term resources
Net Debt to EBITDA	Net Debt/EBITDA	Capacity to finance debt
Capex to Dep.	Capex/Depreciation	Rate at which assets are renewed
Current Assets	Current Assets/Total Assets	Measure of short-term resources
Current Liab.	Current Liab./Total Liab.	Measure of short-term liabilities

Notes: The table gives the description of the financial variables retained. Innovations in these variables are used to control for the impact of financial factors when assessing the predictive contents of ESG ratings on the idiosyncratic volatility of a firm's assets.

4.2 Backtest results

Using the three categories of variables defined above as ESG ratings, idiosyncratic volatility and innovations in financial variables, we compute our test statistics and make inference for the predictive content of the two ESG rating systems considered. To predict the target idiosyncratic volatility variable, we consider a pooled OLS regression for the two models needed to run our backtesting procedure, which are the model that contains only innovations in the $p = 10$ financial variables, and the model that extends this set to include the lagged values of the ESG ratings. Recall that our procedure compare the predictive performance of the two models:

$$\log\left(IRV_{i,t+1}\right) = \alpha_0 + \beta_0 X_{i,t} + \varepsilon^{(0)}_{i,t+1} \tag{19}$$

$$\log\left(IRV_{i,t+1}\right) = \alpha_1 + \beta_1 X_{i,t} + \gamma ESG_{i,t} + \varepsilon^{(1)}_{i,t+1}, \tag{20}$$

where $X_{i,t}$ denotes the vector of innovations in financial variables.

In line with our out-of-sample testing environment, we consider two different forecasting schemes: (i) a fixed forecasting scheme where the first 75% of the total $T = 106$ months for each firm are used to estimate both models, and the forecasts are computed over the last 25% of observations, which are considered as the test sample; (ii) a rolling-window forecasting scheme with the forecasts computed by moving the estimation sample forward by including one more month and excluding the first, giving different estimation samples with the same fixed size of $b = [0.75T]$.

Table 4 displays the outcome of the test for each provider across the three panel datasets. The test statistics are computed using the squared error loss. To gain more insights on the predictive power of the ratings, we perform inference on the aggregate ESG ratings of each providers and also on the specific dimensions of the ratings (environmental, social and governance). The values displayed represent the MSE variation in percentage when the ESG rating (in column) is added to the information set containing only innovations in financial variables. This presentation allows us to test our null hypothesis and to measure the magnitude of the predictive accuracy gain. Negative values are associated to MSE reductions with respect to the model excluding information about the ESG rating (or rating component), and hence to gains in predictive ability. We also report the sign of the regression coefficient associated with the ratings in parentheses. For the rolling window forecasting scheme, the coefficient is averaged across the estimation windows.

For the Europe (EU) universe, the inclusion of ESG information significantly improves the model's predictive accuracy in all configurations except one. Among Sustainalytics

Table 4: Backtest of ESG ratings: results for squared error loss and idiosyncratic returns from CAPM

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−3.0%⋆⋆⋆** **(−0.01)**	**−3.8%⋆⋆⋆** **(−0.009)**	**−1.9%⋆⋆⋆** **(−0.008)**	**−0.54%⋆⋆⋆** **(−0.005)**
	NA	0.12% (−0.01)	**−0.53%⋆⋆⋆** **(−0.008)**	0.51% (−0.006)	0.21% (−0.006)
	AP	**−0.21%⋆⋆** **(−0.004)**	**−0.56%⋆⋆⋆** **(−0.004)**	−0.030% (−0.001)	0.019% (0.0001)
Fixed Window	EU	**−4.0%⋆⋆⋆** **(−0.01)**	**−4.7%⋆⋆⋆** **(−0.009)**	**−2.6%⋆⋆⋆** **(−0.008)**	**−1.1%⋆⋆⋆** **(−0.007)**
	NA	0.81% (−0.01)	−0.42% (−0.01)	1.3% (−0.008)	0.54% (−0.01)
	AP	**−0.55%⋆** **(−0.005)**	**−0.66%⋆⋆** **(−0.004)**	−0.17% (−0.003)	**−0.30%⋆** **(−0.003)**
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−3.1%⋆⋆⋆** **(−0.008)**	**−3.1%⋆⋆⋆** **(−0.006)**	**−3.5%⋆⋆⋆** **(−0.007)**	−0.063% (−0.002)
	NA	−0.0069% (−0.006)	**−0.30%⋆** **(−0.003)**	−0.10% (−0.005)	0.39% (−0.002)
	AP	**−0.028%⋆⋆** **(3e−05)**	**−0.12%⋆⋆⋆** **(−0.0003)**	**−0.36%⋆⋆⋆** **(−0.0008)**	−0.061% (0.0006)
Fixed Window	EU	**−4.2%⋆⋆⋆** **(−0.008)**	**−3.3%⋆⋆⋆** **(−0.005)**	**−4.5%⋆⋆⋆** **(−0.006)**	**−0.55%⋆** **(−0.003)**
	NA	−0.67% (−0.008)	−0.52% (−0.004)	−0.71% (−0.007)	0.51% (−0.003)
	AP	**−0.17%⋆⋆⋆** **(−0.0005)**	**−0.30%⋆⋆⋆** **(−0.0006)**	**−0.50%⋆⋆⋆** **(−0.0009)**	0.081% (−0.0003)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

ratings, the environmental rating is the strongest predictor of idiosyncratic volatility, with an MSE reduction of 3.8% and 4.7% for the rolling and fixed forecasting schemes, respectively. Among Asset4 ratings, the social rating provides more information, resulting in a 3.5% (4.5%) reduction in MSE using a rolling (fixed) forecasting scheme. Overall, the governance rating appears to be less informative in predicting stock return idiosyncratic volatility, as it is associated with the lowest predictive accuracy gains in all configurations. The results are mixed for the North America (NA) and Asia-Pacific (AP) universes. For the NA universe, and for both ESG rating systems, only the inclusion of the environmental rating in the rolling window forecasting scheme leads to significant predictive accuracy gains. The predictive accuracy gains are also lower than those for the EU universe. For the AP universe, we reject our null hypothesis in several configurations, but predictive accuracy gains remain modest compared to those for the EU universe. Furthermore, for most rejections of our null hypothesis, we find a negative association between ESG ratings and idiosyncratic volatility, indicating that higher ESG ratings are, on average, associated with lower stock return idiosyncratic volatility.

4.3 Robustness to factor models

Here we evaluate the sensitivity of our results to the choice of factor model used to compute the target idiosyncratic realized volatility variable. We thus extend the CAPM model and consider a multifactorial model. This extension is anchored to the findings of academic research into the existence of common risk factors beyond the market index. This strand of the literature, which can be dated back to the seminal work of Fama and French (1992), has discovered many market variables or factors that may be able to explain the cross-sectional variations of stock returns. These include the size and value factors in Fama and French (1992) and the momentum factor in Jegadeesh and Titman (1993).

To consider the multifactorial model, we extend the CAPM model in (17) by adding investable factors identified in the literature to drive the cross-sectional variations of the stock's returns. For the Europe and the North America universes these are the MSCI Small/Large Capitalisation factor, which approximates the size anomaly, the MSCI Value/Growth factor associated with the value premium, the MSCI Momentum factor, the MSCI quality factor, and the MSCI Minimum Volatility factor. The lack of data for the Asia-Pacific universe means we consider three factors beyond the market, these being the MSCI Small/Large Capitalisation factor, the MSCI Value/Growth factor, and the MSCI Minimum Volatility factor. Table 5 displays the tests results for the idiosyncratic volatility computed using a multifactorial model and the squared error loss function. Overall, we reach qualitatively

Table 5: Backtest of ESG ratings: results for squared error loss and idiosyncratic returns from multifactorial model

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−3.3%⋆⋆⋆** **(−0.01)**	**−4.2%⋆⋆⋆** **(−0.01)**	**−2.1%⋆⋆⋆** **(−0.009)**	**−0.67%⋆⋆⋆** **(−0.006)**
	NA	−0.076% (−0.01)	**−0.63%⋆⋆⋆** **(−0.009)**	0.39% (−0.009)	0.088% (−0.009)
	AP	**−0.14%⋆** **(−0.003)**	**−0.59%⋆⋆⋆** **(−0.005)**	**−0.030%⋆** **(0.0002)**	0.028% (0.002)
Fixed Window	EU	**−4.4%⋆⋆⋆** **(−0.01)**	**−5.0%⋆⋆⋆** **(−0.01)**	**−2.9%⋆⋆⋆** **(−0.009)**	**−1.3%⋆⋆⋆** **(−0.008)**
	NA	0.56% (−0.02)	−0.43% (−0.01)	1.2% (−0.01)	0.27% (−0.01)
	AP	**−0.42%⋆** **(−0.005)**	**−0.65%⋆⋆** **(−0.005)**	−0.084% (−0.001)	−0.13% (−0.002)
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−3.4%⋆⋆⋆** **(−0.009)**	**−3.4%⋆⋆⋆** **(−0.006)**	**−3.6%⋆⋆⋆** **(−0.007)**	−0.054% (−0.003)
	NA	−0.33% (−0.008)	**−0.53%⋆⋆** **(−0.004)**	−0.26% (−0.007)	0.42% (−0.003)
	AP	0.15% (0.0008)	**−0.043%⋆⋆⋆** **(−9e−05)**	**−0.089%⋆⋆⋆** **(−0.0001)**	**−0.23%⋆⋆⋆** **(0.001)**
Fixed Window	EU	**−4.5%⋆⋆⋆** **(−0.009)**	**−3.6%⋆⋆⋆** **(−0.006)**	**−4.6%⋆⋆⋆** **(−0.007)**	**−0.64%⋆⋆** **(−0.003)**
	NA	**−1.2%⋆** **(−0.01)**	**−0.86%⋆** **(−0.005)**	−0.88% (−0.008)	0.38% (−0.005)
	AP	0.15% (0.0004)	**−0.17%⋆⋆⋆** **(−0.0003)**	**−0.042%⋆⋆⋆** **(−8e−05)**	**−0.11%⋆⋆** **(0.0004)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

similar conclusions, suggesting that our results are robust to the choice of the factor model.

Table 5, which displays the backtest results using the squared loss error, shows similar results as Table 4. For the absolute error loss, results are displayed in appendix B (see Table B.2) and are to be compared to the ones in table B.1. Taken together, these results suggest that our previous conclusions are robust to the choice of the factor model. For the rest of the paper, we restrict our analysis to the dependent variable constructed using the

multifactorial model and to the squared error loss function.

To evaluate the sensitivity of the test to the choice of the loss function, Table B.1 and Table B.2 displays the results using the absolute error loss function for the idiosyncratic volatility from the CAPM and multifactorial model respectively. In comparison with the squared error loss function, the absolute error loss function is more robust to outliers. We find that results are highly similar for the two loss functions, suggesting that our results are robust to the choice of the loss function.

So far, our results show that the predictive power of ESG ratings varies depending on the universe considered. We find strong evidence that higher ESG ratings are associated with lower future stock return idiosyncratic volatility for the European universe, and to a lesser extent for the North America and Asia-Pacific universes. This finding can be explained by the fact that European regulation on ESG issues is more stringent, with the establishment of a high-level expert group on sustainable finance (HLEG) in 2016 and the subsequent introduction of the EU taxonomy for sustainable activities.[6] As a result, European investors are more likely to consider ESG information valuable for their investment decisions compared to US investors (Amel-Zadeh and Serafeim, 2018). Regarding the rating dimensions, the environmental rating appears to carry the most information, followed by the social rating, while predictive accuracy gains are consistently lower for the governance rating. This is consistent with the findings of Berg et al. (2020), who reported that the noise in ratings is higher for the governance component, followed by the social component, with the environmental component being the least noisy. In the next subsection, we conduct additional empirical investigations to check the robustness of our results.

4.4 Disagreement between raters and the informational content of the ESG ratings

Our results suggest that both rating systems are informative for forecasting idiosyncratic volatility in Europe, where regulation on ESG is more stringent, and to a lesser extent in other regions. Another factor that could affect the link between ESG ratings and return volatility is ESG ratings disagreement. Serafeim and Yoon (2021) analyzed the link between ESG ratings and ESG risks as measured by ESG-related events and showed that the consensus rating predicts future news, but its predictive ability diminishes for firms where there is a large disagreement between raters. They also found that the consensus rating moderates the stock market reaction to ESG risks. Therefore, the forecasting power of ESG ratings could be moderated by ESG rating disagreement as it affects both the likelihood of ESG events and the stock market reaction to ESG risk materialization. In our sample, we

[6] https://finance.ec.europa.eu/publications/high-level-expert-group-sustainable-finance-hleg_en

find that ESG ratings are quite divergent across the three universes. The R-squared for the linear regression between the two rating agencies is equal to 40.88% for the EU universe, 46.46% for the NA universe, and 32.65% for the AP universe (see Figures 4, B.3, and B.4).

Table 6: Distribution of sectors in consensus and disagreement samples

Sector	Consensus	Disagreement
Consumer Discretionary	20.3%	19.5%
Industrials	20.3%	21.1%
Information Technology	14.7%	13.5%
Energy	10.2%	8.4%
Materials	9.1%	12.0%
Consumer Staples	8.1%	8.0%
Healthcare	6.1%	6.0%
Communication Services	5.6%	5.2%
Utilities	4.6%	5.6%
Financials	1.0%	0.8%

Notes: The table displays distribution of sectors in the consensus and disagreement samples.

To check for this stylised fact, we replicate the results of Table 5 but partition each panel into consensus and disagreement groups, based on the firm level correlation between the ratings of the two providers. For each universe, the consensus group contains firms belonging to the top 25% of highest correlations, while the disagreement group contains firms belonging to the top 25% lowest correlations. Among the consensus group, the average correlation between ESG ratings of the two providers are equal to 75%, 72% and 70% for the EU, NA and AP universes, respectively. Among the disagreement group, these figures are equal to -35%, -46% and -44%, meaning that there is considerable divergence between ratings. Table 6 shows the sector distribution in the consensus and disagreement samples. Since the consensus rating is driven by the different methodologies used by rating agencies rather than by firm characteristics (Berg et al., 2020), we observe a similar distribution of sectors across the two groups.

Table 7 displays the backtest results for the consensus and disagreement groups using a rolling window forecasting scheme. The results using a fixed window are displayed in Appendix B (Table B.3). We observe significant differences in terms of the rejection of the null hypothesis between the two groups. Among the consensus group, we observe 22 rejections out of 24 tests at the 1% nominal risk level, while this figure drops to 15 rejections for the disagreement group. Moreover, the forecasting power of ESG ratings is consistently greater for consensus firms across the three universes. For example, considering the EU universe, the MSE reduction due to the inclusion of Sustainalytics environmental rating is

Table 7: Consensus vs disagreement between providers using a rolling window forecasting scheme (MSE)

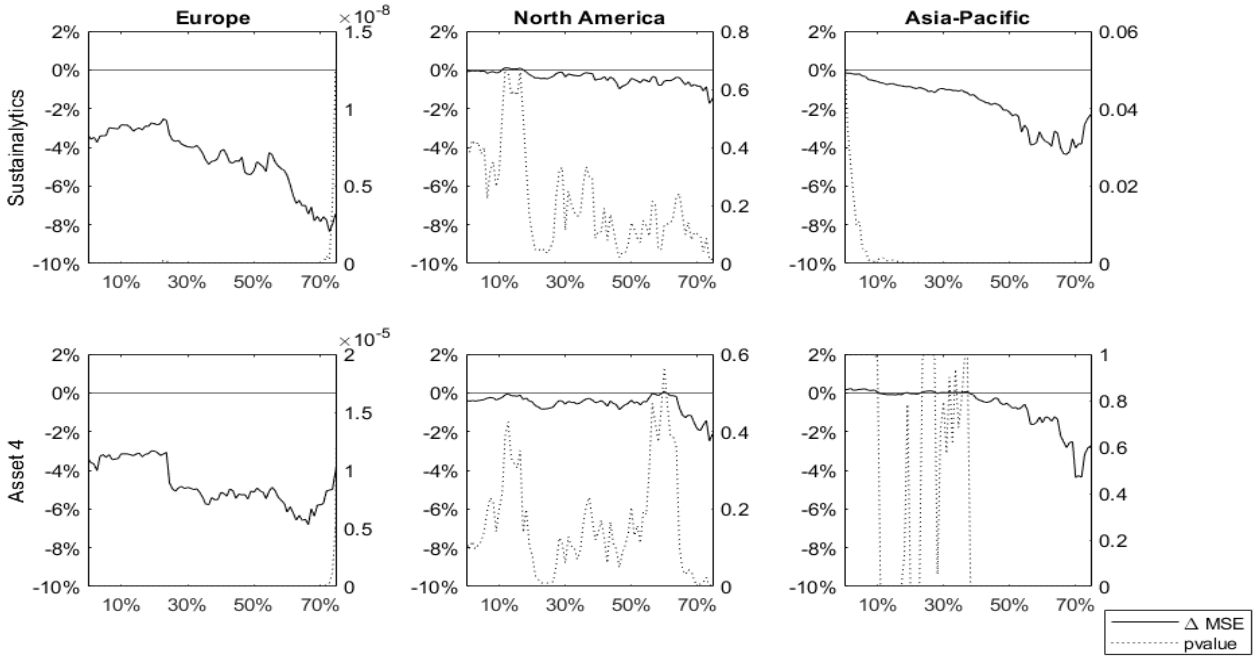
		Sustainalytics			
		ESG	E	S	G
Consensus	EU	**−7.7%***** **(−0.02)**	**−7.1%***** **(−0.01)**	**−5.0%***** **(−0.01)**	**−3.7%***** **(−0.01)**
	NA	**−1.6%***** **(−0.02)**	**−2.6%***** **(−0.01)**	**−1.0%***** **(−0.01)**	0.97% (−0.009)
	AP	**−2.4%***** **(−0.01)**	**−3.5%***** **(−0.01)**	**−0.64%***** **(−0.003)**	**−1.1%***** **(−0.008)**
Disagreement	EU	**−2.9%***** **(−0.01)**	**−2.3%***** **(−0.007)**	**−2.8%***** **(−0.01)**	**−0.032%***** **(−0.003)**
	NA	**−0.091%***** **(−0.01)**	**−0.14%***** **(−0.01)**	**0.035%**** **(−0.009)**	**0.027%**** **(0.0003)**
	AP	**−1.3%***** **(0.006)**	0.29% (−0.001)	**−1.1%***** **(0.009)**	**−1.4%***** **(0.008)**
		Asset 4			
		ESG	E	S	G
Consensus	EU	**−4.0%***** **(−0.008)**	**−4.8%***** **(−0.007)**	**−4.7%***** **(−0.007)**	**−0.056%***** **(−0.0002)**
	NA	**−2.3%***** **(−0.009)**	**−5.1%***** **(−0.008)**	**−2.1%***** **(−0.006)**	0.46% (−0.002)
	AP	**−2.5%***** **(−0.006)**	**−3.2%***** **(−0.005)**	**−0.78%***** **(−0.003)**	**−0.33%***** **(−0.001)**
Disagreement	EU	**−1.2%***** **(−0.008)**	**−3.4%***** **(−0.005)**	**−1.4%***** **(−0.008)**	0.74% (−0.001)
	NA	0.64% (−0.008)	0.21% (−0.003)	0.55% (−0.005)	0.48% (−0.005)
	AP	**−0.30%***** **(0.002)**	**−0.92%***** **(0.002)**	0.28% (0.001)	**−0.24%***** **(0.0007)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a rolling window forecasting scheme. For a given universe, the consensus group contains firms with the 25% highest correlations between the ratings of the two providers. The disagreement group contains firms with the 25% lowest correlations. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

equal to 7.7% in the consensus sample, but only 2.3% in the disagreement sample. For the NA universe, this decrease reaches 2.6% for the consensus sample and 0.14% for the disagreement sample. Similar conclusions hold for most configurations and universes.

To assess the sensitivity of the previous results to the threshold used to define the consensus firms, we repeated the analysis for alternative levels of ESG consensus. We started with the full sample and excluded the top x% of firms with the highest level of disagreement before applying our inferential procedure. Figure 8 displays the results for values of x ranging between 0% and 75% using a rolling window forecasting scheme. Results obtained using a fixed window forecasting scheme are displayed in the appendix (Figure B.9). We find that the forecasting power of ESG ratings increases with the level of ESG consensus. This result is consistent for both rating agencies and across the three universes. Overall, predictive accuracy gains due to the inclusion of ESG information increase with the level of ESG consensus.

Figure 8: Decrease in forecast error in function of ESG consensus (rolling window)



Source: This table displays the variation in MSE when ESG information is included in the model as a function of the level of consensus between ESG providers. The x-axis represents the level of consensus between rating agencies. For a level of consensus x, only the firms with the $1 - x$ highest correlations between the ratings of the two providers were included in the sample.

We next test if the predictive ability of ESG ratings is the same for consensus firms with high and low ratings. To do so, we apply our test separately to consensus firms with a high ESG rating (above the median) and a low ESG rating (below the median). Results using

a rolling (fixed) window forecasting scheme are displayed in Table 8 (Table B.4). We find that for both rating agencies, the predictive ability is greater for consensus firms with a low rating in the NA universe, but that the predictive accuracy gains depend on the rating agency considered for the other universes.

Table 8: Consensus firms: high vs low ESG rating (rolling window)

		Sustainalytics			
		ESG	E	S	G
High ESG	EU	**−2.9%⋆⋆⋆** **(−0.01)**	**−5.8%⋆⋆⋆** **(−0.01)**	−0.16% (−0.005)	0.19% (0.002)
	NA	4.0% (0.01)	0.67% (0.002)	6.8% (0.009)	−0.37% (−0.003)
	AP	**−0.68%⋆⋆** **(−0.005)**	−0.47% (−0.005)	**−0.75%⋆⋆** **(−0.003)**	0.037% (6e−05)
Low ESG	EU	**−4.8%⋆⋆⋆** **(−0.01)**	**−4.3%⋆⋆⋆** **(−0.009)**	**−2.1%⋆⋆⋆** **(−0.008)**	**−3.7%⋆⋆⋆** **(−0.01)**
	NA	**−1.7%⋆⋆** **(−0.02)**	**−2.6%⋆⋆⋆** **(−0.01)**	−0.60% (−0.01)	0.12% (−0.008)
	AP	**−1.9%⋆⋆⋆** **(−0.005)**	**−0.63%⋆⋆** **(−0.0007)**	**−1.4%⋆⋆⋆** **(−0.005)**	**−2.1%⋆⋆⋆** **(−0.008)**
		Asset 4			
		ESG	E	S	G
High ESG	EU	**−8.5%⋆⋆⋆** **(−0.02)**	**−2.5%⋆⋆⋆** **(−0.005)**	**−6.6%⋆⋆⋆** **(−0.01)**	**−1.2%⋆⋆⋆** **(−0.003)**
	NA	0.23% (0.003)	−0.75% (−0.004)	0.66% (0.008)	0.88% (−0.002)
	AP	**−3.6%⋆⋆⋆** **(−0.009)**	**−0.97%⋆⋆** **(−0.004)**	**−0.86%⋆⋆** **(−0.003)**	**−0.54%⋆⋆** **(−0.0009)**
Low ESG	EU	0.25% (0.0007)	**−2.8%⋆⋆⋆** **(−0.004)**	**−0.71%⋆⋆⋆** **(−0.001)**	**−1.7%⋆⋆⋆** **(0.005)**
	NA	**−4.6%⋆⋆⋆** **(−0.01)**	**−7.0%⋆⋆⋆** **(−0.009)**	**−4.3%⋆⋆⋆** **(−0.008)**	**−0.20%⋆** **(−0.0003)**
	AP	−0.14% (−0.0006)	−0.074% (−0.0003)	−0.11% (0.0008)	0.47% (−0.001)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model for consensus firms. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a rolling window forecasting scheme. High (low) ESG sample represents firms above (below) the median ESG rating. ⋆, ⋆⋆ and ⋆⋆⋆ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

From a practical point of view, our results provide crucial information for portfolio managers who integrate ESG information into their investment decisions. We show that it is necessary to cross-check the information gathered from multiple ESG rating providers

before integrating ESG into the management process. The focal point of our results is that consensus about the ESG ratings is informative about idiosyncratic risk, while ESG ratings with disagreement are less valuable from this viewpoint.

5 Conclusion

The contribution of this article is to propose a formal statistical procedure for assessing the informational content in ESG ratings. The test proceeds by evaluating how well these extra-financial metrics help in predicting a given target variable intended to measure firm-specific risks. Our framework allows users to choose a target variable related to their investment objectives. Technically, our inferential procedure for checking the informational content in ESG ratings is based on extending the conditional predictive ability test of Giacomini and White (2006) to a panel setting. Under weak assumptions, including cross-sectional dependencies among loss functions for firms, we derive the Gaussian asymptotic distribution of the test statistic. Monte Carlo simulations conducted under different types of model misspecification show that the test has good small sample properties.

Empirical applications are conducted using the idiosyncratic volatility of stock returns, a measure of firm-specific risk, as our target variable. We apply our procedure to evaluate two leading ESG rating systems (Sustainalytics and Asset4) in three investment universes (Europe, North America, and the Asia-Pacific region). The results show that the null hypothesis of a lack of informational content in ESG ratings is strongly rejected for Europe, while the results are mixed and predictive accuracy gains are lower for the other regions. Furthermore, we find that the predictive accuracy gains are higher for the environmental dimension of the ESG ratings. Importantly, we find that the predictive accuracy gains derived from ESG ratings increase with the level of consensus between rating agencies in all three universes, while they are low for firms over which there is a high level of disagreement.

The results have important implications for investors and researchers. For investors, our backtest procedure provides a useful and practical framework for considering ESG rating providers before integrating the ratings into the investment process. Our results suggest prudence about the information content of ESG ratings when they diverge. For researchers in asset pricing, it is crucial to check properly the quality of ESG ratings before using them, especially when the ratings are divergent. Moreover, the link between ESG ratings and idiosyncratic volatility when the ratings are convergent suggests that ESG investing is not just an issue of the preferences of investors, but that ESG ratings can also provide information about future fundamentals and risks. A future application for investors could be to compare the ratings of competing ESG rating agencies, since our inferential procedure

can be easily adapted to compare the informational content in the ESG ratings. This would help investors in selecting one agency among several competing ones in non-nested comparisons, or in considering additional competing agencies to combine with their already existing ratings in nested comparisons.

References

O. Akgun, A. Pirotte, G. Urga, and Z. Yang. Equal predictive ability tests for panel data with an application to oecd and imf forecasts. Unpublished Manuscript, 2020.

R. Albuquerque, Y. Koskinen, and Z. Chendi. Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science*, 65:4451–4469, 2019.

A. Amel-Zadeh and G. Serafeim. Why and how investors use esg information: Evidence from a global survey. *Financial Analyst Journal*, 74(3):87–103, 2018.

D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

D. Ardia, K. Bluteau, K. Boudt, and K. Inghelbrecht. Climate change concerns and the performance of green vs. brown stocks. *Management Science*, 2022.

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

F. Berg, J. Koelbel, and R. Rigobon. Aggregate confusion: the divergence of esg ratings. *Massachusetts Institute of Technology*, 2020.

F. Berg, J. F. Koelbel, A. Pavlova, and R. Rigobon. Esg confusion and stock returns: Tackling the problem of noise. Technical report, National Bureau of Economic Research, 2022.

M. Billio, M. Costola, I. Hristova, C. Latino, and L. Pelizzon. Inside the esg ratings: (dis)agreement and performance. *Unpublished Manuscript*, 2019.

K. Bouslah, L. Kryzanowski, and B. M'Zali. The impact of the dimensions of social performance on firm risk. *Journal of Banking and Finance*, 37:1258–1273, 2013.

C. Champagne, F. Coggins, and A. Sodjahin. The performance of extra-financial ratings as measure of esg-risk. Unpublished Manuscript, 2019.

A. Chatterji, D. Levine, and M. Toffel. How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy*, 18(1):125–169, 2009.

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

A. Davies and K. Lahiri. A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, 68:205–228, 1995.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263, 1995.

E. Dimson, P. Marsh, and M. Staunton. Divergent esg ratings. *The Journal of Portfolio Management*, 47(1):75–87, 2020.

A. Dyck, V. Lins, L. Roth, and H. Wagner. Do institutional investors drive corporate social responsibility? *Journal of Financial Economics*, 131(3):693–714, 2019.

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

E. F. Fama and K. R. French. Disagreement, tastes, and asset prices. *Journal of Financial Economics*, 83(3):667–689, 2007.

R. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica*, 74(6): 1545–1578, 2006.

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

C. Gollier and S. Pouget. Investment strategies and corporate behaviour with socially responsible investors: A theory of active ownership. *Economica*, 89(356):997–1023, 2022.

S. Hartzmark and A. Sussman. Do investors value sustainability? a natural experiment examining ranking and fund ows. *Journal of Finance*, 74(6):2789–2837, 2019.

A. Hoepner, I. Oikonomou, Z. Sautner, L. Starks, and X. Zhou. Esg shareholder engagement and downside risk. *Unpublished Manuscript*, 2018.

E. Ilhan, Z. Sautner, and G. Vilkov. Carbon tail risk. *The Review of Financial Studies*, 34 (3):1540–1571, 2019.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993.

H. Jo and H. Na. Does csr reduce firm risk? evidence from controversial industry sectors. *Journal of business ethics*, 110:441–456, 2012.

S. Mishra and S. B. Modi. Positive and negative corporate social responsibility, financial leverage, and idiosyncratic risk. *Journal of business ethics*, 117:431–448, 2013.

K. Mozaffar, G. Serafeim, and A. Yoon. Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6):1697–1724, 2016.

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

L. Pedersen, S. Fitzgibbons, and L. Pomorski. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, 2020.

A. Riedl and P. Smeets. Why do investors hold socially responsible mutual funds. *Journal of Finance*, 72(6):2505–2550, 2017.

N. Semenova and L. Hassel. On the validity of environmental performance metrics. *Journal of Business Ethics*, 132(2):249–258, 2015.

G. Serafeim and A. Yoon. Stock price reactions to esg news: The role of esg ratings and disagreement. *Working paper 21-079, Harvard Business School*, 2021.

G. Serafeim and A. Yoon. Which corporate esg news does the market react to? *Financial Analysts Journal*, 78(1):59–78, 2022.

A. Sodjahnin, C. Champagne, F. Coggins, and R. Gillet. Leading or lagging indicators of risk? the informational content of extra-financial performance scores. *Journal of Asset Management*, 18(5):347–370, 2017.

A. Timmermann and Y. Zhu. Comparing forecasting performance with panel data. Unpublished Manuscript, 2019.

K. D. West. Asymptotic inference about predictive ability. *Econometrica*, 64:1067–1084, 1996.

H. White. *Asymptotic Theory for Econometricians*. Academic Press, 2001. Revised Edition.

O. D. Zerbib. A sustainable capital asset pricing model (s-capm): Evidence from environmental integration and sin stock exclusion. *Review of Finance*, 26(6):1345–1388, 2022.

A Appendix A: Details on the Monte Carlo simulations

In this Appendix we provide details about the simulations of innovations in the financial variables for generating the small sample properties of the test (see Section 3). These variables are generated via a multivariate Gaussian distribution with mean vector $\bar{x}$ and covariance matrix Ω calibrated using real data. The dataset we use contains historical monthly values of $p = 10$ innovations in the financial variables for 238 European firms from January 2010 to October 2018, giving a total of 106 months.

Innovations are computed as deviations from the overall means. The financial variables are, in order: tax burden ratio, interest burden ratio, operating margin ratio, asset turnover ratio, leverage as measured by the ratio of total assets to total equity, current ratio as measured by the ratio of current assets to current liabilities, debt ratio, capex as measured by the ratio of capital expenditures to depreciation, current assets as measured by the ratio of current assets to total assets, current liabilities as measured by the ratio of current liabilities to total liabilities.

The mean vector is thus equal to

$$\bar{x} = [0.8137; 0.8333; 0.1391; 0.8265; 3.8713; 1.4031; 1.7466; 1.2779; 0.3634; 0.2880],$$

and the covariance matrix Ω equal to

$$\Omega = \begin{pmatrix} 4.098 & -0.061 & -0.007 & -0.003 & -0.003 & 0.008 & 0.209 & -0.023 & -0.001 & -0.001 \\ -0.061 & 17.732 & -0.008 & 0.037 & 5.704 & 0.057 & -0.136 & -0.021 & 0.017 & 0.007 \\ -0.007 & -0.008 & 0.012 & -0.025 & -0.369 & 0.010 & -0.018 & 0.025 & -0.005 & -0.006 \\ -0.003 & 0.037 & -0.025 & 0.284 & -0.559 & -0.025 & -0.358 & -0.067 & 0.042 & 0.037 \\ -0.003 & 5.704 & -0.369 & -0.559 & 5012.291 & -0.521 & 30.792 & 4.391 & -0.160 & -0.092 \\ 0.008 & 0.057 & 0.010 & -0.025 & -0.521 & 0.642 & -0.420 & 0.042 & 0.053 & -0.040 \\ 0.209 & -0.136 & -0.018 & -0.358 & 30.792 & -0.420 & 32.172 & 0.550 & -0.154 & -0.058 \\ -0.023 & -0.021 & 0.025 & -0.067 & 4.391 & 0.042 & 0.550 & 1.841 & -0.023 & -0.022 \\ -0.001 & 0.017 & -0.005 & 0.042 & -0.160 & 0.053 & -0.154 & -0.023 & 0.029 & 0.014 \\ -0.001 & 0.007 & -0.006 & 0.037 & -0.092 & -0.040 & -0.058 & -0.022 & 0.014 & 0.018 \end{pmatrix}.$$

For the simulation of the target variable of idiosyncratic volatility, we run a pooled OLS regression with the dependent variable being the logarithm of the monthly time series of idiosyncratic realised volatility over the same period (January 2010 to October 2018) for the 238 European firms. The explanatory variables are the innovations in the 10 financial variables as described above.

$c^\star$	$\beta_1^\star$	$\beta_2^\star$	$\beta_3^\star$	$\beta_4^\star$	$\beta_5^\star$	$\beta_6^\star$	$\beta_7^\star$	$\beta_8^\star$	$\beta_9^\star$	$\beta_{10}^\star$
-5.9165	0.0070	-0.0015	-0.8739	-0.0679	0.0048×10^{-2}	0.0941	0.0044	0.0605	0.1869	0.0824

For the $p = 10$ financial variables, the estimated coefficients are displayed above. These estimates are used to generate data for simulating the logarithm of idiosyncratic realised volatility, and applying the exponential function leads to the target variable.

B Appendix B: Additional Tables and Figures

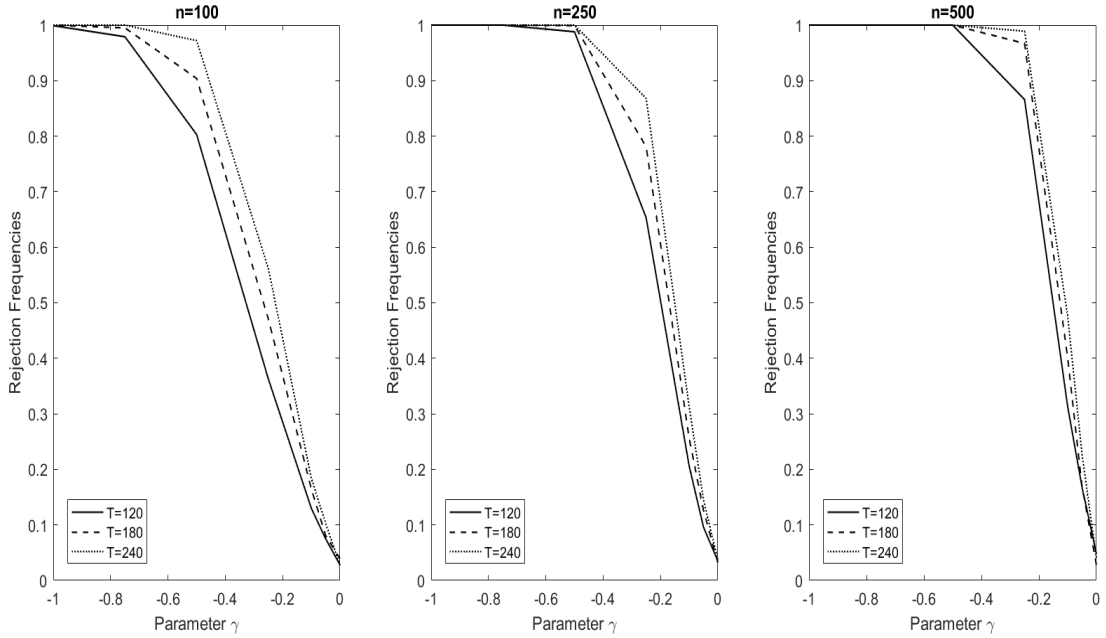


Figure B.1: Rejection Frequencies under a medium level of misspecification with the absolute error loss function

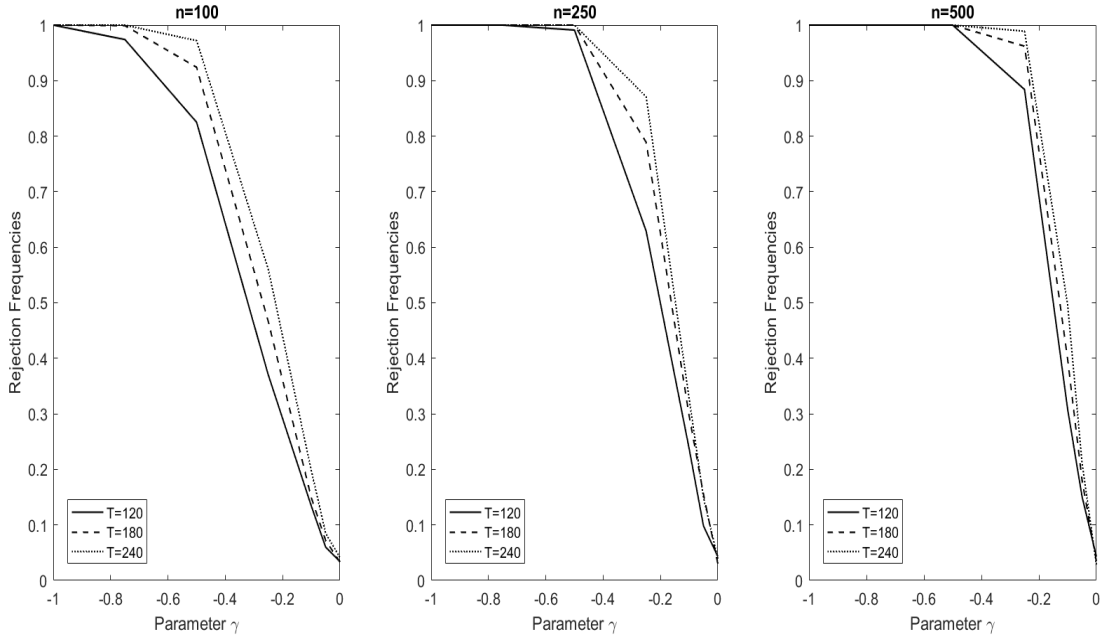
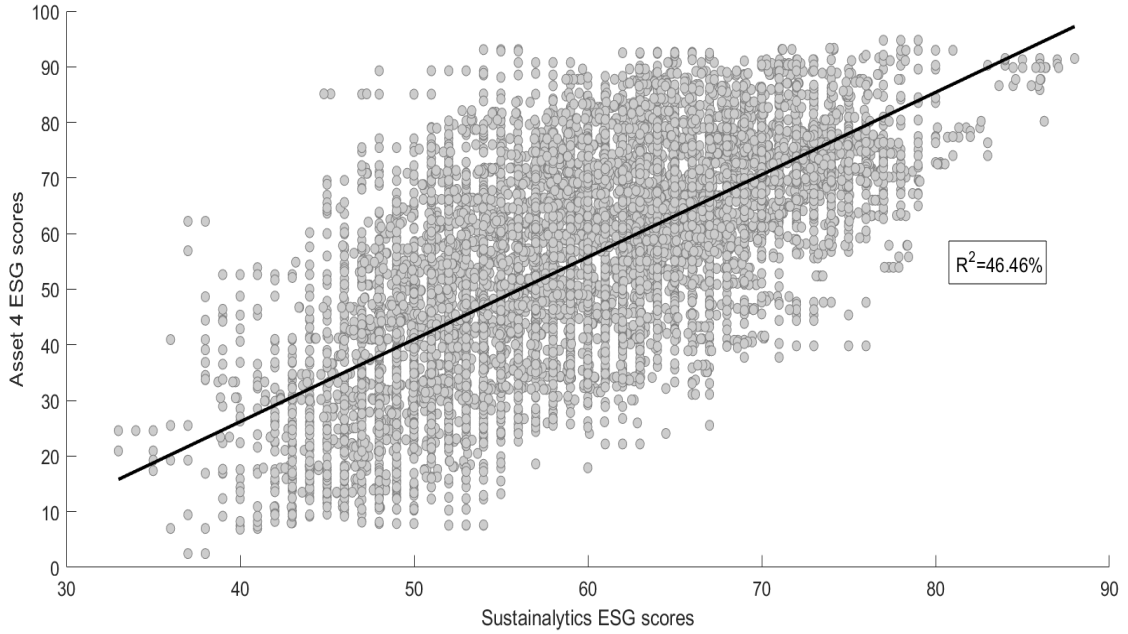


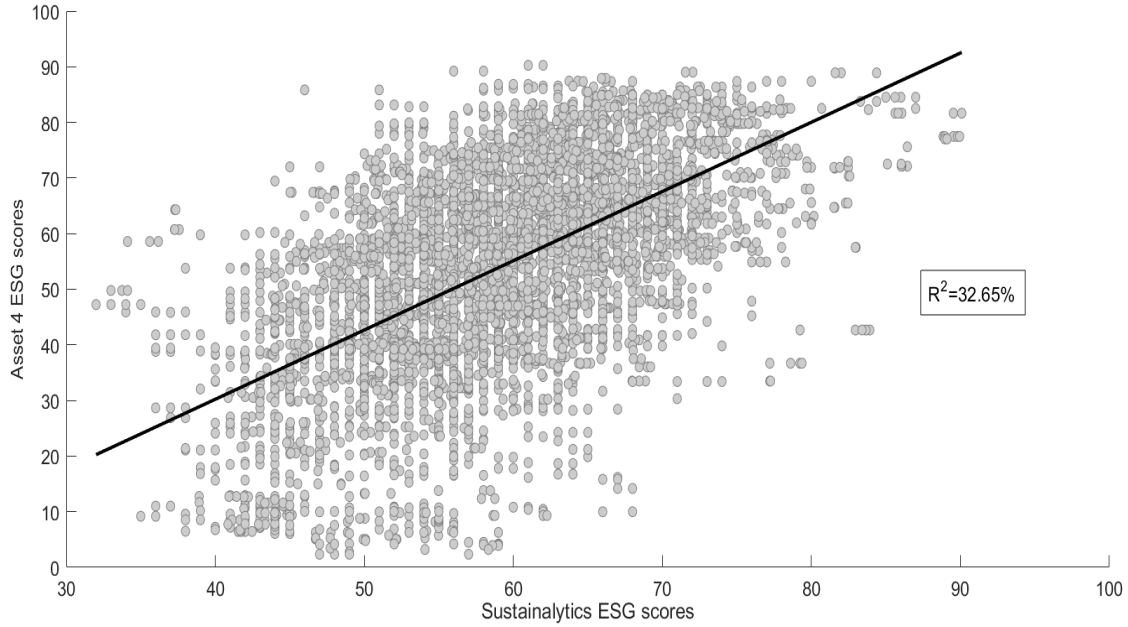
Figure B.2: Rejection Frequencies under a high level of misspecification with the absolute error loss function

Figure B.3: Relation between the Sustainalytics and Asset4 ESG ratings: North America



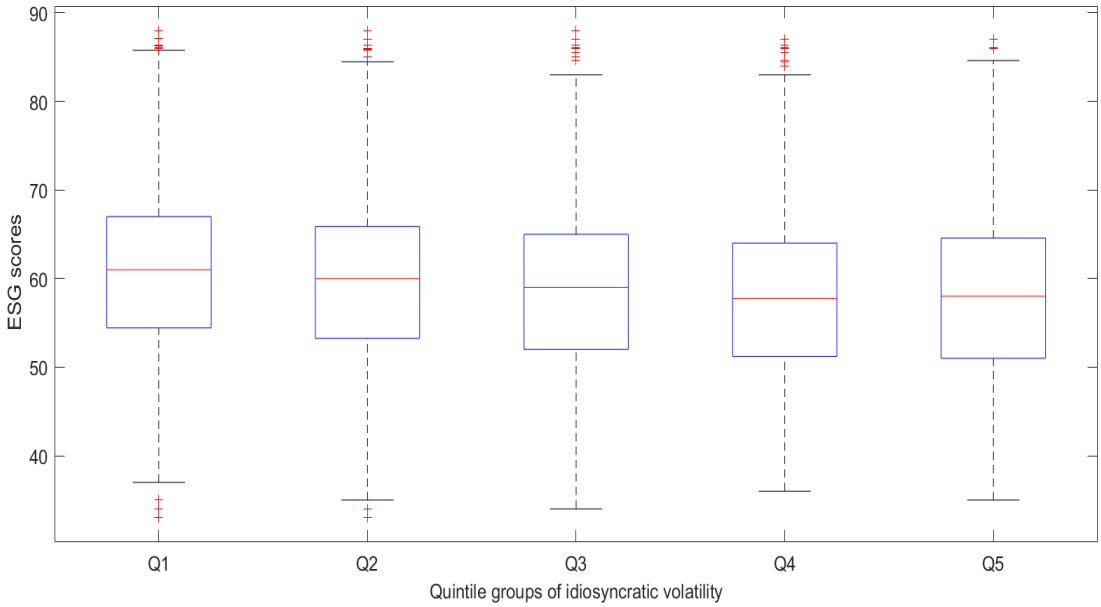
Source: The figure displays the scatter plot that shows the graphical relation between the ESG ratings for the two providers considered (Sustainalytics and Asset4). The datasets contain monthly observations for $n = 326$ firms from January 2010 to October 2018, giving a total of 106 months.

Figure B.4: Relation between the Sustainalytics and Asset4 ESG ratings: Asia-Pacific



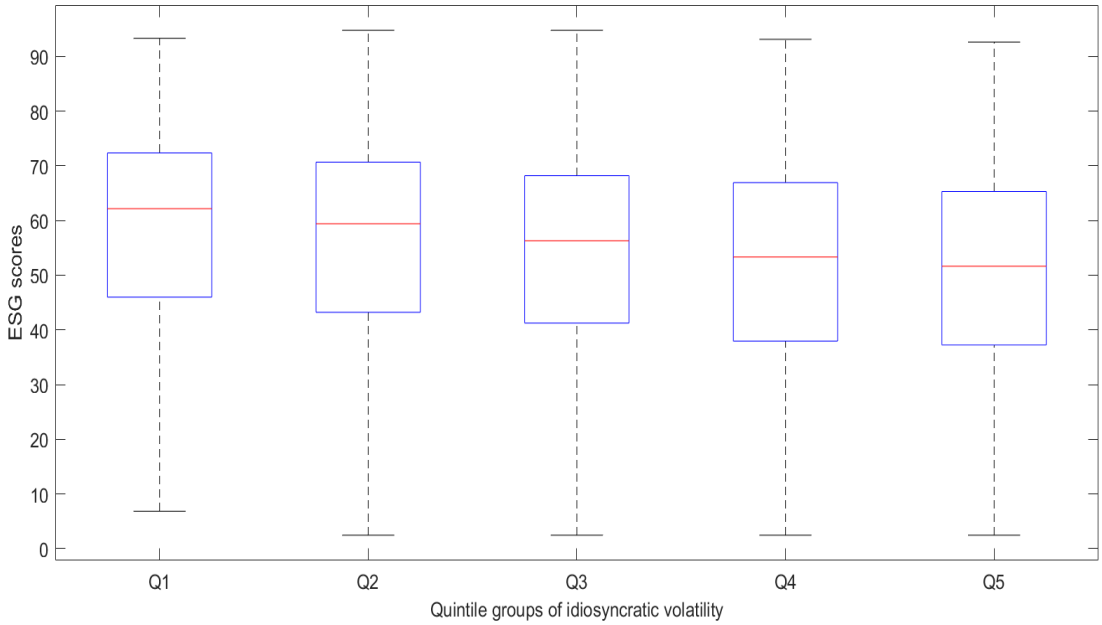
Source: The figure displays the scatter plot that shows the graphical relation between the ESG ratings for the two providers considered (Sustainalytics and Asset4). The datasets contain monthly observations for $n = 217$ firms from January 2010 to October 2018, giving a total of 106 months.

Figure B.5: ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (North America)



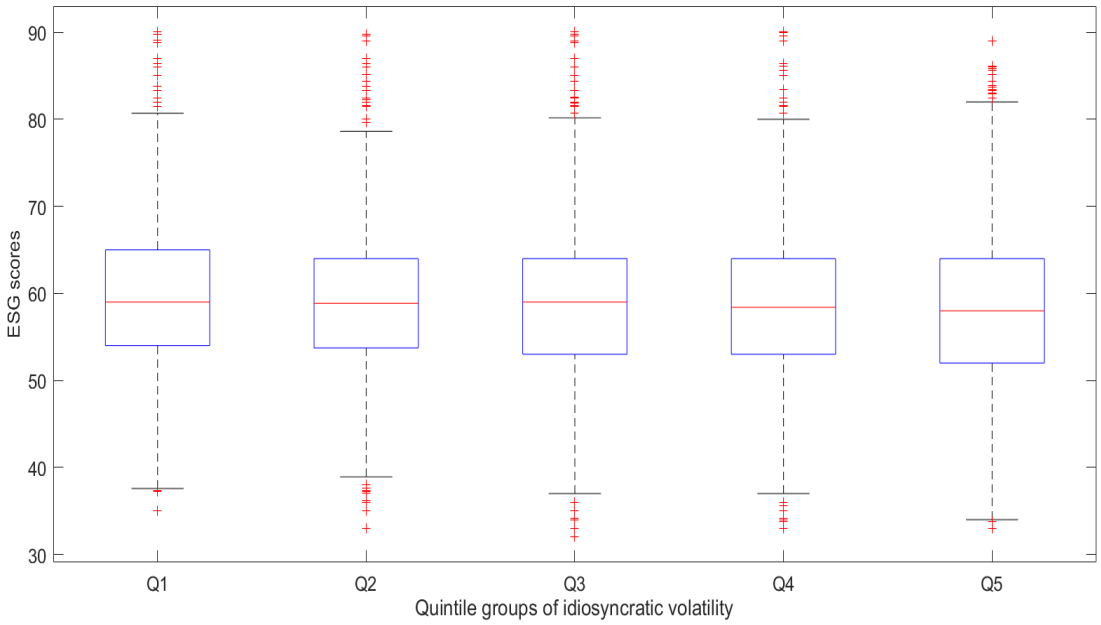
Source: For the North America universe, the figure displays the means of Sustainalytics ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 326$ firms from January 2010 to October 2018, giving a total of 106 months.

Figure B.6: ESG ratings by idiosyncratic volatility quintiles: Asset4 (North America)



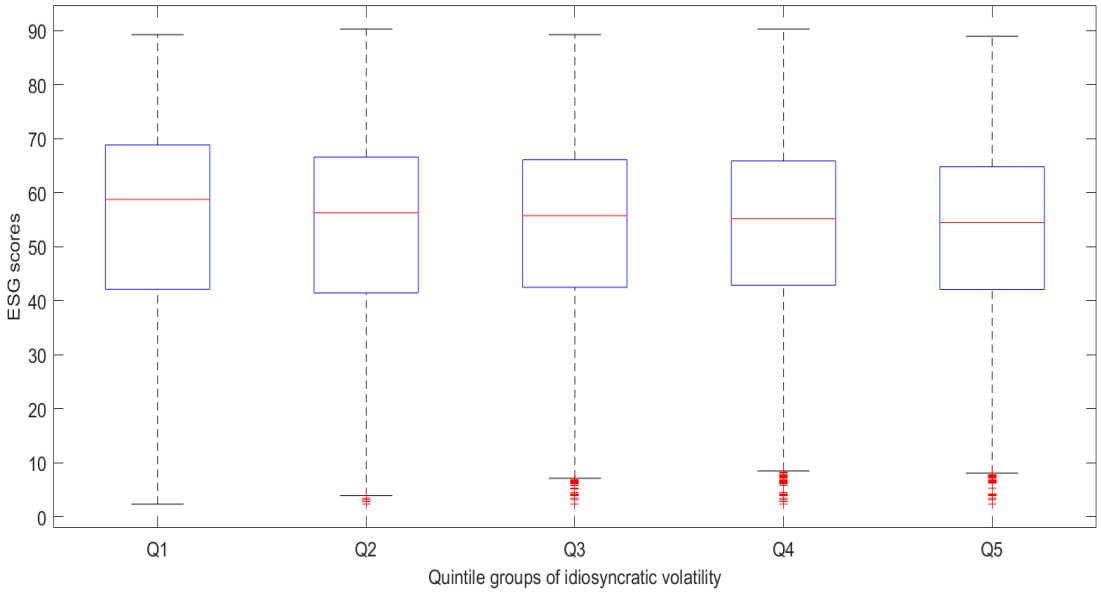
Source: For the North America universe, the figure displays the means of Asset4 ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 326$ firms from January 2010 to October 2018, giving a total of 106 months.

Figure B.7: ESG ratings by idiosyncratic volatility quintiles: Sustainalytics (Asia-Pacific)



Source: For the Asia-Pacific universe, the figure displays the means of Sustainalytics ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 217$ firms from January 2010 to October 2018, giving a total of 106 months.

Figure B.8: ESG ratings by idiosyncratic volatility quintiles: Asset4 (Asia-Pacific)



Source: For the Asia-Pacific universe, the figure displays the means of Asset4 ESG ratings within the five groups defined by the quintiles of idiosyncratic volatility computed with residual asset returns from the CAPM. The dataset contains monthly observations for $n = 217$ firms from January 2010 to October 2018, giving a total of 106 months.

Table B.1: Backtest of ESG ratings: results for absolute error loss and idiosyncratic returns from CAPM

		Sustainalytics			
		ESG	E	S	G
Rolling Window	EU	**$-1.6\%^{\star\star\star}$** **(−0.01)**	**$-2.1\%^{\star\star\star}$** **(−0.009)**	**$-0.98\%^{\star\star\star}$** **(−0.008)**	**$-0.38\%^{\star\star\star}$** **(−0.005)**
	NA	-0.075% (−0.01)	**$-0.43\%^{\star\star\star}$** **(−0.008)**	0.17% (−0.006)	0.060% (−0.006)
	AP	**$-0.14\%^{\star\star}$** **(−0.004)**	**$-0.39\%^{\star\star\star}$** **(−0.004)**	-0.0024% (−0.001)	0.0079% (0.0001)
Fixed Window	EU	**$-2.3\%^{\star\star\star}$** **(−0.01)**	**$-2.7\%^{\star\star\star}$** **(−0.009)**	**$-1.4\%^{\star\star\star}$** **(−0.008)**	**$-0.74\%^{\star\star\star}$** **(−0.007)**
	NA	0.14% (−0.01)	**$-0.56\%^{\star\star}$** **(−0.01)**	0.52% (−0.008)	0.22% (−0.01)
	AP	**$-0.37\%^{\star\star}$** **(−0.005)**	**$-0.44\%^{\star\star}$** **(−0.004)**	-0.12% (−0.003)	**$-0.19\%^{\star\star}$** **(−0.003)**
		Asset 4			
		ESG	E	S	G
Rolling Window	EU	**$-2.0\%^{\star\star\star}$** **(−0.008)**	**$-1.6\%^{\star\star\star}$** **(−0.006)**	**$-2.3\%^{\star\star\star}$** **(−0.007)**	**$-0.19\%^{\star\star\star}$** **(−0.002)**
	NA	**$-0.23\%^{\star}$** **(−0.006)**	**$-0.26\%^{\star\star}$** **(−0.003)**	**$-0.29\%^{\star\star}$** **(−0.005)**	0.10% (−0.002)
	AP	$2.9e-05\%$ (3e−05)	**$-0.069\%^{\star\star\star}$** **(−0.0003)**	**$-0.17\%^{\star\star\star}$** **(−0.0008)**	**$-0.062\%^{\star\star}$** **(0.0006)**
Fixed Window	EU	**$-2.5\%^{\star\star\star}$** **(−0.008)**	**$-1.8\%^{\star\star\star}$** **(−0.005)**	**$-2.8\%^{\star\star\star}$** **(−0.006)**	**$-0.46\%^{\star\star\star}$** **(−0.003)**
	NA	**$-0.81\%^{\star\star}$** **(−0.008)**	**$-0.51\%^{\star\star}$** **(−0.004)**	**$-0.76\%^{\star\star}$** **(−0.007)**	0.12% (−0.003)
	AP	**$-0.089\%^{\star\star\star}$** **(−0.0005)**	**$-0.18\%^{\star\star\star}$** **(−0.0006)**	**$-0.25\%^{\star\star\star}$** **(−0.0009)**	0.055% (−0.0003)

Notes: This table displays the variation in mean absolute error (MAE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from CAPM. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Table B.2: Backtest of ESG ratings: results for absolute error loss and idiosyncratic returns from multifactorial model

Sustainalytics					
		ESG	E	S	G
Rolling Window	EU	**−1.8%⋆⋆⋆** **(−0.01)**	**−2.3%⋆⋆⋆** **(−0.01)**	**−1.0%⋆⋆⋆** **(−0.009)**	**−0.44%⋆⋆⋆** **(−0.006)**
	NA	−0.16% (−0.01)	**−0.52%⋆⋆⋆** **(−0.009)**	0.16% (−0.009)	0.037% (−0.009)
	AP	**−0.11%⋆⋆** **(−0.003)**	**−0.45%⋆⋆⋆** **(−0.005)**	0.0095% (0.0002)	0.033% (0.002)
Fixed Window	EU	**−2.3%⋆⋆⋆** **(−0.01)**	**−2.7%⋆⋆⋆** **(−0.01)**	**−1.4%⋆⋆⋆** **(−0.009)**	**−0.81%⋆⋆⋆** **(−0.008)**
	NA	0.034% (−0.02)	**−0.56%⋆⋆** **(−0.01)**	0.49% (−0.01)	0.028% (−0.01)
	AP	**−0.34%⋆⋆** **(−0.005)**	**−0.49%⋆⋆** **(−0.005)**	−0.066% (−0.001)	**−0.090%⋆** **(−0.002)**
Asset 4					
		ESG	E	S	G
Rolling Window	EU	**−2.0%⋆⋆⋆** **(−0.009)**	**−1.7%⋆⋆⋆** **(−0.006)**	**−2.2%⋆⋆⋆** **(−0.007)**	**−0.20%⋆⋆** **(−0.003)**
	NA	**−0.31%⋆** **(−0.008)**	**−0.34%⋆⋆** **(−0.004)**	**−0.29%⋆** **(−0.007)**	0.17% (−0.003)
	AP	0.097% (0.0008)	**−0.022%⋆⋆** **(−9e−05)**	**−0.026%⋆⋆** **(−0.0001)**	**−0.14%⋆⋆⋆** **(0.001)**
Fixed Window	EU	**−2.5%⋆⋆⋆** **(−0.009)**	**−1.8%⋆⋆⋆** **(−0.006)**	**−2.7%⋆⋆⋆** **(−0.007)**	**−0.52%⋆⋆⋆** **(−0.003)**
	NA	**−0.86%⋆⋆** **(−0.01)**	**−0.57%⋆⋆** **(−0.005)**	**−0.68%⋆** **(−0.008)**	0.064% (−0.005)
	AP	0.084% (0.0004)	**−0.11%⋆⋆⋆** **(−0.0003)**	**−0.024%⋆⋆⋆** **(−8e−05)**	**−0.065%⋆⋆** **(0.0004)**

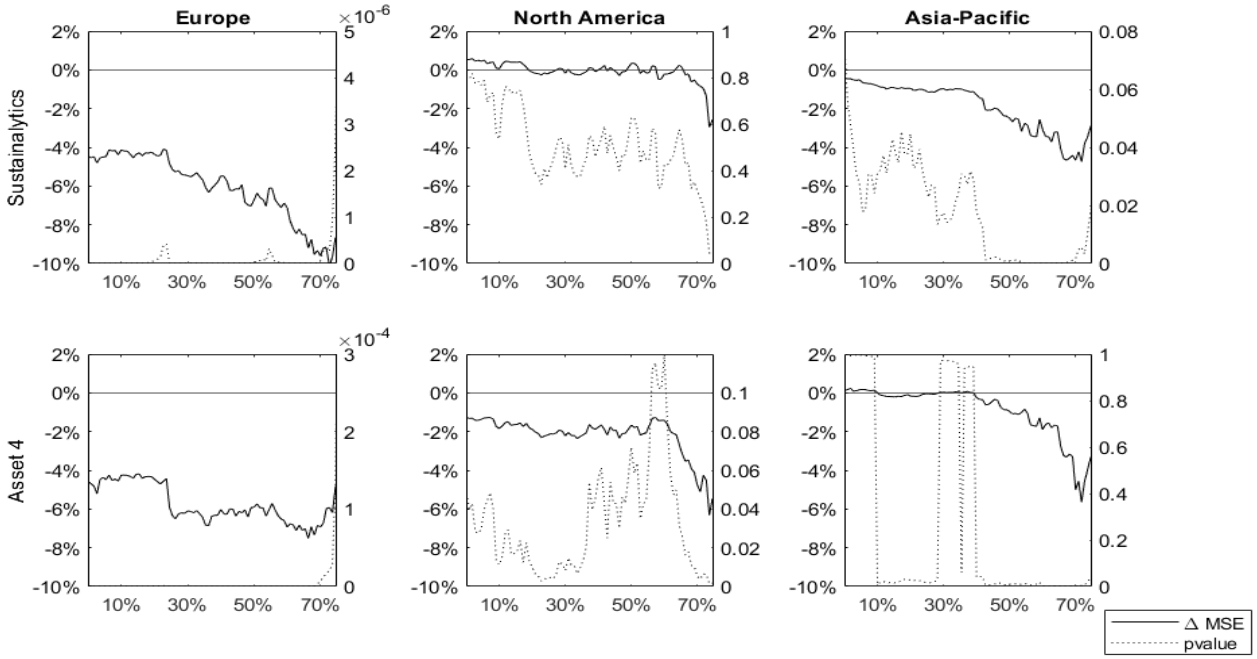
Notes: This table displays the variation in mean absolute error (MAE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. The datasets contain monthly observations from January 2010 to October 2018, giving a total of 106 months. The North America, Europe and Asia-Pacific datasets include information on respectively $n = 326$, $n = 238$ and $n = 217$ firms. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Table B.3: Consensus vs disagreement between providers using a fixed window forecasting scheme (MSE)

		Sustainalytics			
		ESG	E	S	G
Consensus	EU	**−8.9%⋆⋆⋆** **(−0.02)**	**−6.8%⋆⋆⋆** **(−0.009)**	**−6.3%⋆⋆⋆** **(−0.01)**	**−5.9%⋆⋆⋆** **(−0.01)**
	NA	**−3.0%⋆⋆⋆** **(−0.02)**	**−4.8%⋆⋆⋆** **(−0.01)**	**−0.93%⋆⋆⋆** **(−0.01)**	0.70% (−0.01)
	AP	**−3.0%⋆⋆⋆** **(−0.01)**	**−2.8%⋆⋆⋆** **(−0.008)**	**−0.49%⋆⋆⋆** **(−0.002)**	**−4.4%⋆⋆⋆** **(−0.01)**
Disagreement	EU	**−1.8%⋆⋆⋆** **(−0.007)**	**−1.8%⋆⋆⋆** **(−0.005)**	**−1.4%⋆⋆⋆** **(−0.005)**	**−0.28%⋆⋆⋆** **(−0.004)**
	NA	**−4.2%⋆⋆⋆** **(−0.02)**	**−4.7%⋆⋆⋆** **(−0.01)**	**−2.5%⋆⋆⋆** **(−0.01)**	**−0.97%⋆⋆⋆** **(−0.004)**
	AP	**−0.63%⋆⋆⋆** **(0.004)**	0.47% (−0.002)	**−0.43%⋆⋆⋆** **(0.006)**	**−0.76%⋆⋆⋆** **(0.003)**
		Asset 4			
		ESG	E	S	G
Consensus	EU	**−4.8%⋆⋆⋆** **(−0.007)**	**−5.1%⋆⋆⋆** **(−0.006)**	**−5.8%⋆⋆⋆** **(−0.006)**	**−0.052%⋆⋆⋆** **(−0.0007)**
	NA	**−6.2%⋆⋆⋆** **(−0.01)**	**−8.3%⋆⋆⋆** **(−0.009)**	**−4.7%⋆⋆⋆** **(−0.007)**	**−0.24%⋆⋆⋆** **(−0.004)**
	AP	**−3.1%⋆⋆⋆** **(−0.007)**	**−3.9%⋆⋆⋆** **(−0.005)**	**−1.2%⋆⋆⋆** **(−0.004)**	**−0.51%⋆⋆⋆** **(−0.002)**
Disagreement	EU	**−1.8%⋆⋆⋆** **(−0.005)**	**−1.8%⋆⋆⋆** **(−0.002)**	**−2.5%⋆⋆⋆** **(−0.006)**	0.84% (−0.003)
	NA	**−5.0%⋆⋆⋆** **(−0.01)**	**−3.0%⋆⋆⋆** **(−0.004)**	**−3.6%⋆⋆⋆** **(−0.008)**	**−3.3%⋆⋆⋆** **(−0.007)**
	AP	0.24% (0.002)	**−0.38%⋆⋆⋆** **(0.002)**	0.58% (0.001)	**−0.059%⋆⋆⋆** **(0.0001)**

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a fixed window forecasting scheme. For a given universe, the consensus group contains firms with the 25% highest correlations between the ratings of the two providers. The disagreement group contains firms with the 25% lowest correlations. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.

Figure B.9: Decrease in forecast error in function ESG consensus (fixed window)



Source: This table displays the variation in MSE when ESG information is included in the model as a function of the level of consensus between ESG providers. The x-axis represents the level of consensus between rating agencies. For a level of consensus x, only the firms with the $1 - x$ highest correlations between the ratings of the two providers were included in the sample.

Table B.4: Consensus firms: high vs low ESG rating (fixed window)

		Sustainalytics			
		ESG	E	S	G
High ESG	EU	−3.2% (−0.01)	**−7.1%⋆⋆⋆ (−0.009)**	−1.7% (−0.01)	3.4% (−0.007)
	NA	2.7% (−0.003)	0.66% (−0.0008)	3.8% (−0.003)	0.91% (−0.003)
	AP	**−1.2%⋆⋆ (−0.005)**	**−2.6%⋆⋆ (−0.009)**	1.2% (0.006)	**−2.4%⋆ (−0.01)**
Low ESG	EU	**−3.6%⋆⋆⋆ (−0.008)**	**−1.6%⋆⋆⋆ (−0.002)**	**−1.7%⋆⋆ (−0.005)**	**−4.9%⋆⋆⋆ (−0.01)**
	NA	**−3.1%⋆⋆ (−0.02)**	**−5.7%⋆⋆⋆ (−0.01)**	−0.59% (−0.02)	0.70% (−0.01)
	AP	−1.8% (−0.01)	−1.4% (−0.006)	0.71% (−0.007)	**−5.2%⋆⋆⋆ (−0.01)**
		Asset 4			
		ESG	E	S	G
High ESG	EU	−3.3% (−0.01)	**−1.4%⋆ (−0.004)**	**−8.6%⋆⋆⋆ (−0.01)**	0.044% (−0.0002)
	NA	15.% (−0.009)	18.% (−0.01)	**−1.5%⋆ (0.002)**	13.% (−0.004)
	AP	**−9.4%⋆⋆⋆ (−0.01)**	**−6.6%⋆⋆⋆ (−0.01)**	**−2.3%⋆⋆⋆ (−0.003)**	0.14% (0.0002)
Low ESG	EU	1.0% (0.002)	**−1.6%⋆⋆⋆ (−0.003)**	0.29% (0.0003)	−1.4% (0.004)
	NA	**−7.6%⋆⋆⋆ (−0.01)**	**−9.7%⋆⋆⋆ (−0.009)**	**−6.3%⋆⋆⋆ (−0.008)**	−0.75% (−0.002)
	AP	0.50% (−0.004)	−1.2% (−0.002)	0.45% (−0.001)	1.0% (−0.002)

Notes: This table displays the variation in mean squared error (MSE) when ESG information is included in the model for consensus firms. Idiosyncratic volatilities are computed using the residual asset returns from a multifactorial model. Results are computed using a fixed window forecasting scheme. High (low) ESG sample represents firms above (below) the median ESG rating. $\star$, $\star\star$ and $\star\star\star$ indicate rejection of the null hypothesis of lack of informational content in ESG ratings at the 10%, 5% and 1% nominal risk levels respectively. Regression coefficients associated to the ESG rating are reported in parentheses.