NBER WORKING PAPER SERIES

ESG CONFUSION AND STOCK RETURNS: TACKLING THE PROBLEM OF NOISE

Florian Berg
Julian F. Koelbel
Anna Pavlova
Roberto Rigobon

Working Paper 30562
http://www.nber.org/papers/w30562

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2022

Florian Berg, Julian Kölbel, and Roberto Rigobon are grateful to Massachusetts Pension Reserves Investment Management Board, AQR Capital Management (Applied Quantitative Research), Qontigo, Asset One, and MFS Investment Management — members of the Aggregate Confusion Project Council— for their generous support of this research. We are also extremely grateful to the ESG rating agencies that provided their data to this project. We thank Ing-Haw Cheng, Aaron Pancost, and Jonathan Parker and seminar participants at the 2022 American Finance Association meetings, ARCS 2022, the 2022 ASU Sonoran Winter Finance Conference, George Washington University, London Business School, MIT GCFP, MIT Sloan, the 2021 PCOB colloquium on ESG Disclosure, the University of Amsterdam, the University of Mannheim, and the University of Michigan for excellent feedback. Julian serves as a member of the RepRisk Academic Advisory Council. All remaining errors are ours. Correspondence to: Roberto Rigobon, Sloan School of Management, MIT, 50 Memorial Drive, E62-520, Cambridge, MA 02142-1347, aggregateconfusion@mit.edu, tel: (617) 258 8374. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

ESG Confusion and Stock Returns: Tackling the Problem of Noise
Florian Berg, Julian F. Koelbel, Anna Pavlova, and Roberto Rigobon
NBER Working Paper No. 30562
October 2022
JEL No. C26,G12,Q56

ABSTRACT

How does ESG (environmental, social, and governance) performance affect stock returns? Answering this question is difficult because existing measures of ESG perfor- mance — ESG ratings — are noisy and, therefore, standard regression estimates suffer from attenuation bias. To address the bias, we propose two noise-correction procedures, in which we instrument ESG ratings with ratings of other ESG rating agencies, as in the classical errors-in-variables problem. The corrected estimates demonstrate that the effect of ESG performance on stock returns is stronger than previously estimated: after correcting for attenuation bias, the coefficients increase on average by a factor of 2.6, implying an average noise-to-signal ratio of 61.7%. The attenuation bias is stable across horizons at which stock returns are measured. In simulations, our noise-correction pro- cedures outperform the standard approaches followed by practitioners such as averages or principal component analysis.

Florian Berg
MIT Sloan School of Management
E62-520100 Main Street
fberg@mit.edu

Julian F. Koelbel
University of St. Gallen
PLD-G 02
Plattenstr. 32
St. Gallen 9000 Switzerland
julian.koelbel@unisg.ch

Anna Pavlova
London Business School
Regents Park
London NW1 4SA
United Kingdom
apavlova@london.edu

Roberto Rigobon
MIT Sloan School of Management
100 Main Street, E62-516
Cambridge, MA 02142
and NBER
rigobon@mit.edu

1 Introduction

ESG (environmental, social, and governance) investing has taken the asset management industry by storm. In the U.S., assets under management of ESG funds increased by 35% in 2021 alone, while in Europe assets in ESG funds increased by 52 billion EUR to hit 1.1 trillion EUR.[1] Unless demand curves for stocks are perfectly elastic, such unprecedented demand for assets with superior ESG performance should boost their prices. While theoretical studies suggest such an effect,[2] the empirical evidence is mixed.[3]

We believe that one of the key reasons confounding the relationship between ESG attributes and stock returns is noise, or put differently, that available measures of ESG performance are noisy. ESG ratings play a crucial role in measuring a firm's ESG attributes, guiding the investment of ESG funds, and thus linking investor preferences for ESG to portfolio choices. These third-party assessments are provided by ESG rating agencies as a commercial service to investors. However, there is significant disagreement between ESG ratings from different providers because each ESG rating is generated by a unique methodology. Methodologies differ due to different ways of choosing and aggregating ESG attributes and different ways of measuring ESG attributes. For example, the average pairwise correlation of the ESG ratings in our sample is only 0.2.[4] Berg, Kölbel, and Rigobon (2020) show that the main source of this disagreement is differences in measurement, which hints that ESG attributes are measured imperfectly. Our goal is to disentangle signal from noise in ESG ratings and to uncover the true impact of ESG performance on expected stock returns.

To do this, we propose a simple model that establishes a relationship between ESG performance and stock returns and show that the noisier the measurement of ESG performance, the lower the sensitivity of stock returns to ESG performance. Moreover, we show that the latter result implies that regression estimates of the relationship between stock returns and noisy measures of ESG performance would be biased towards zero; in other words, the noisier the measurement, the larger the bias. To address the bias, we develop two noise-correction procedures. Specifically, we instrument a given ESG rating agency's score with other rating agencies' scores as in the classical errors-in-variables problem. We conduct this analysis for the eight largest ESG rating agencies, for firms located in the eurozone, Japan, the U.K.,

[1] See `https://assets.contentstack.io/v3/assets/blt4eb669caa7dc65b2/blt7a208fcfc3d719a8/61ade16b7de7d945b9c4b8cd/European_ESG_Fund_Landscape_2020.pdf`, accessed August 17, 2022.

[2] Heinkel, Kraus, and Zechner (2001); Pastor, Stambaugh, and Taylor (2021b); Fama and French (2007)

[3] A recent meta-study by Atz, Bruno, Liu, and Van Holt (2022) identifies 1141 peer-reviewed academic papers written between 2015 and 2020 that investigate the link between ESG and financial performance.

[4] The ESG rating agencies in our dataset include ISS ESG (majority stake owned by Deutsche Boerse), MSCI IVA (owned by MSCI), RepRisk (independent), Refinitiv (formerly known as Asset4), SP Global CSA (formerly known as RobecoSAM), Sustainalytics (owned by Morningstar), Truvalue Labs (owned by FactSet), and Moody's (formerly known as Vigeo-Eiris). See Section 2 for more details.

and the U.S. The corrected estimates demonstrate that the effect of ESG performance on stock returns is stronger than previously estimated: after correcting for attenuation bias, the coefficients increase on average by a factor of 2.6. The attenuation bias is stable over different time horizons over which we measure returns. We determine which agencies' scores are valid instruments (not all of them are) and estimate the noise-to-signal ratio for each ESG rating agency (some of which are very large).

In the theoretical part of this paper, we propose a simple model that establishes a relationship between ESG performance and stock returns and show that the noisier the measurement of ESG performance, the lower the sensitivity of stock returns to ESG performance. In our model are two types of investors: traditional and ESG-conscious. The former care only about a firm's cash flow, whereas the latter care additionally about the ESG performance of their portfolio holdings. The non-pecuniary ESG attribute is uncorrelated with the firm's cash flow. We assume further that information about this ESG attribute is contained in a noisy ESG signal, provided by a rating agency. We derive stock prices in closed form and conclude that the noisier the ESG signal, the lower its effect on stock prices. This dampening effect is similar to the attenuation bias arising in OLS regressions. The regression of interest in our case is the regression of stock returns on ESG scores, which are noisy signals of true (unobserved) ESG performance.

In the empirical part of this paper, we use an instrumental variable approach to address the measurement error problem and correct the attenuation bias. Specifically, we propose to instrument a rating of one agency by the ratings of other agencies for the same attribute. Standard regressions then need to be replaced by two-stage least squares (2SLS) regressions. We document that the effect of ESG performance on stock returns is much stronger — the coefficients on average more than double — when we replace the standard OLS regression by 2SLS. This result is consistent with the prediction of the theory that the bias we see in the standard regressions is indeed an attenuation bias. This could well be the reason why many studies do not observe an effect of ESG performance on stock returns.

We propose two procedures for selecting ESG ratings as instruments in the 2SLS estimation. While each rating agency's ESG score is well predicted by a combination of other rating agencies' scores, this is not sufficient to guarantee that ESG scores are valid instruments. To test for instrument validity, we conduct overidentifying restrictions (OIR) tests. As we have a total of 8 ESG rating agencies in our sample, we can test multiple OIR. In our first procedure, which we term *Pruning*, we chose the instruments by starting from the largest possible set and pruning instruments one at a time until the model passes the Sargan-Hansen test of OIR. The failure of the test indicates that some instruments are invalid. The main problem with this procedure is that the Sargan-Hansen test has not been designed for this sequential search. Our second procedure, which we call *Lasso*, starts with a minimal set of instruments and adds instruments as long as the OIR test is not being rejected. For both procedures, we find that many but not all of the ESG ratings pass the OIR tests. This can happen if ESG scores used as instruments are, for example, backfilled retroactively by a provider or if scores of one ESG rater are influenced by another. Another issue could be that measurement errors are correlated across rating agencies (because agencies use similar procedures or rely on imputed data to arrive at the scores). Our OIR tests diagnose these violations and we exclude scores that are invalid instruments from our estimation.

We argue theoretically that the attenuation bias, captured by the ratio between the 2SLS and OLS coefficients, should be invariant to the horizon over which stock returns are measured. This is indeed confirmed by our empirical analysis. We estimate the model for 1-month through 8-month returns and find that the ratios between the two coefficients are statistically indistinguishable from each other. Furthermore, the ratio between the 2SLS coefficient and its OLS counterpart can be estimated for each individual rating agency and geographical region. This ratio is a measure of the implied noise in the rating agency's score and we find in our empirical section that the average noise-to-signal ratio is 61.7% across regions, raters, and horizons. We also show that ESG ratings from different rating agencies have different levels of noise and document large variation in noise-to-signal ratios across regions.

It is important to highlight that all raters' scores are valuable. Disregarding scores of some raters amounts to discarding valuable information about the imperfectly measured ESG attributes. By combining information from several complementary ratings, one can obtain a more precise estimate of the impact of ESG performance on stock returns. It is also important to mention that our procedure does not produce or estimate a less noisy ESG score. Strictly speaking, our procedure only solves the problem of noise in estimating the relationship between ESG performance and stock returns. The impact of noise on the relationship between ESG performance and other variables (e.g., accounting measures of performance) is not addressed here, but might be evaluated using the same method.

We run simulations to compare our noise-correction procedures to common alternative approaches such as a simple average or principal component analysis. The simulations show that our procedure performs significantly better than the alternatives. Suppose, for example, that one ESG rating is noisier than another. A simple average ignores this information and puts the same weights on the two ratings. The principle component analysis is designed to explain observed variance. It would therefore put the largest weight on the signal with the highest variance and most likely the highest noise. An ideal approach should instead put the lowest weight on the noisiest variable, and this is what our procedure effectively does.

In additional simulations, we focus on potential model misspecifications and find that the OIR test has significant power in our setting. Finally, we recognize that in practice ESG ratings are aggregates of multiple indicators (e.g., carbon emissions, labor practices, etc.), and raters choose different sets of indicators in constructing their scores. This means that we have fewer instruments than possible sources of noise. In simulations, however, we show that our procedure still goes a long way in recovering the effect of ESG performance on stock returns, i.e., the 2SLS estimates are much closer to the true coefficient than their OLS counterparts.

Our paper is related to several strands of literature. First, it is related to asset pricing models that incorporate ESG investors who push up asset prices of green firms and reduce their cost of capital (Heinkel, Kraus, and Zechner, 2001; Friedman and Heinle, 2016; Oehmke and Opp, 2019; Broccardo, Hart, and Zingales, 2022; Landier and Lovo, 2020; Kashyap, Kovrijnykh, Li, and Pavlova, 2021; Pastor, Stambaugh, and Taylor, 2021b). In the same vein is also Pedersen, Fitzgibbons, and Pomorski (2021). Our study shows that these theoretical predictions are hard to detect empirically due to noisy measurement of ESG performance.

Second, numerous studies have explored the link between ESG performance and stock returns empirically. However, the evidence is not conclusive; studies report both higher stock returns for ESG performers (Edmans, 2011; Khan, Serafeim, and Yoon, 2016; Lins, Servaes, and Tamayo, 2017; Albuquerque, Koskinen, and Zhang, 2019) as well as lower stock returns (Chava, 2014; El Ghoul, Guedhami, Kwok, and Mishra, 2011; Bolton and Kacperczyk, 2020). Pastor, Stambaugh, and Taylor (2021a) stress the importance of distinguishing between expected and realized stock returns, and argue that the expected stock returns of high ESG performers is lower. While our model is fully consistent with Pastor, Stambaugh, and Taylor, we add an important point that regardless of whether one focuses on expected or realized returns, noisy measurement will tend to attenuate the effect.

Third, our paper is related to empirical studies in finance and accounting that have explored the relationship between corporate governance and stock returns (Gompers, Ishii, and Metrick, 2003; Bauer, Guenster, and Otten, 2004; Adams and Ferreira, 2009; Bebchuk, Cohen, and Ferrell, 2009). This literature has addressed the problem of how corporate governance ought to be measured by suggesting several alternative approaches to measurement (e.g. Larcker, Richardson, and Tuna, 2007; Daines, Gow, and Larcker, 2010; Larcker, Reiss, and Xiao, 2015). Our instrumental variable approach offers an innovative way to examine the relationship between corporate governance and stock returns when there are competing ways of measuring it.

Finally, our paper is related to the literature on ESG rating divergence (Berg, Kölbel, and Rigobon, 2020; Christensen, Hail, and Leuz, 2021; Christensen, Serafeim, and Sikochi, 2022). Two recent papers that study the consequences of ESG rating divergence at the firm level, Avramov, Cheng, Lioui, and Tarelli (2021) and Gibson, Krueger, and Schmidt (2021), suggest that uncertainty about ESG performance leads to a higher risk premium. Our perspective is different. We interpret ESG rating divergence as measurement error, which attenuates the true effect of ESG performance on stock returns in standard regressions.

2 Data

2.1 ESG Ratings

ESG rating agencies offer a commercial service to investors by providing third-party assessments of firms' ESG performance. Different ESG raters provide diverging ESG ratings, as the correlations in Table 3 confirm. This is because each ESG rating is generated by a unique methodology. Methodologies differ due to different ways of measuring ESG attributes and different ways of choosing and aggregating ESG attributes. Berg, Kölbel, and Rigobon (2020) show that measurement is the main source of divergence, followed by the choice, and then the aggregation of attributes.

ESG rating agencies resort to a variety of data sources for their assessment. A key challenge is that there is only a limited amount of standardized and publicly available data about companies' ESG performance. Mainly, data comes from five distinct sources, namely from companies' own ESG reports, regulatory filings, the media, questionnaires that rating agencies send to companies, and modelled data. For example, some rating agencies model carbon emissions to make up for missing data on carbon emissions. These sources differ along important dimensions, namely whether the information is available to the public or not, who reports the information (the company itself or a third-party observer), whether disclosure is mandatory or voluntary, and whether it follows disclosure standards such as the Global Reporting Initiative (GRI) or the Sustainability Accounting Standards Board (SASB). As a result, the indicators that ESG ratings are built upon are noisy.

ESG rating agencies determine which attributes should be evaluated as part of their scoring procedure and how important they are relative to each other. The list of relevant attributes typically includes attributes such as green house emissions, product safety, or labor practices, but can also include less obvious attributes such as electromagnetic radiation, management of systemic risks, or whether top management has monetary incentives to meet ESG targets. The weight of these attributes can also differ, and in many cases weights are industry specific and determined according to a proprietary methodology. ESG raters attempt to aggregate ESG attributes in a way that is consistent with what a representative

ESG investor cares about.[5] As a result, the way that ESG ratings are produced implies that each of them offers a noisy measurement of some underlying true ESG performance, which itself remains unobservable.

For example, at an individual indicator level (e.g., CO_2 emissions), "true" means precisely the actual CO_2 emissions that occurred. The noise is the difference between what the rating agency uses for CO_2 emissions and the actual one. For ESG ratings, which are themselves weighted averages of indicators, "true" means that the indicators are measurement-error free and that the weights assigned to the indicators coincide with the weights that the representative ESG investor assigns to individual ESG attributes.

The ESG rating agencies included in our dataset are shown in Table 1. The table also shows the alternative names or ownership as well as the exact name of the scores used in the analysis. We use ISS's Numeric ESG Overall Rating, Moody's Global score, MSCI's IVA Industry Weighted score, Refinitiv's TRESG score, RepRisk's Reputation Risk Index (RRI), Sustainalytics' ESG Risk Rating, S&P Global's ESG score, and the Insight Score from Truvalue Labs (TVL).

Some ratings have changed their methodology during the study period. For example, MSCI updated its methodology in 2017. More problematically, some raters may have retroactively changed their scores, which has been shown in the case of Refinitiv (Berg, Fabisik, and Sautner, 2021). Sustainalytics has provided us with a dataset that is simulated backwards from 2018, based on their new methodology. It is possible that ESG rating data that are not point-in-time have a potential look-ahead bias. We address this issue as part of our methodology.

Table 1. ESG Scores Overview. This table shows the current and previous names of the data providers, ownership, and the exact name of the scores used in the analysis.

Rater Name	Previous Name	Owner	Score Name
ISS ESG	Oekom Research	ISS Inc	Numeric ESG Overall Rating
Moody's	Vigeo-Eiris	Moody's	Global Score
MSCI	Innovest	MSCI Inc.	IVA Industry Weighted Score
Refinitiv	Asset4	London Stock Exchange Group	TRESG Score
RepRisk	–	RepRisk AG	Reputation Risk Index (RRI)
Sustainalytics	Sustainalytics	Morningstar	ESG Risk Rating
S&P Global CSA	RobecoSAM	S&P Global	ESG Score
Truvalue Labs	–	FactSet	Insight Score

[5] See McCahery, Sautner, and Starks (2016) for an analysis of the corporate governance preferences of different institutional investors.

2.2 Financial Data

Financial data comes from Compustat's Capital IQ. All data for the eurozone, the U.K., Japan, and the U.S. is in local currency. *Return* is the return data expressed in percentage points. *Beta* is the market beta estimated from monthly returns from month -60 to month -1. *Dividends* are the dividends per share over the prior 12 months divided by price at the end of the prior month. *Market Value* is the logarithm of the market value of equity at the end of the prior month. *Book-to-market* is the logarithm of book equity minus the logarithm of market value of equity at the end of the prior month. *Asset Growth* is the logarithm of growth in total assets in the prior fiscal year. *ROA* is the income before extraordinary items divided by average total assets in the prior fiscal year. *Momentum* is the return from month -12 to month -2. *Volatility* is the monthly standard deviation, estimated from daily returns from month -12 to month -1. All financial variables are winsorized at the 1% level.

2.3 Descriptives

Table 2 presents the descriptive statistics for the ESG variables as well as the financial variables. Refinitiv, RepRisk, S&P Global, Sustainalytics, Truvalue Labs, as well as Moody's have a rating on a scale from 0 to 100, ISS from 1 to 4, and MSCI from 0 to 10. We multiply RepRisk's and Sustainalytics' scores by -1 and add 100 to align them with the other ratings. A high value of a rating signifies a good performance and a low rating a bad performance. The analysis is performed over the time period from January 2015 to December 2020, the starting point being determined by Sustainalytics' data, which starts in December 2014. The sample consists of 273 firms and 15531 firm-month observations for the eurozone, 131 firms and 6655 firm-month observations for the U.K., 246 firms and 16147 firm-month observations for Japan, and 506 firms and 26974 firm-month observations for the U.S.

The descriptive statistics of the financial variables in the U.S. are in line with Lewellen (2015)'s large stocks sample. The average market value is 9.1 billion EUR, 5.3 billion GBP, 900 billion JPY, and 18.2 billion USD for the eurozone, U.K., Japan, and U.S., respectively. This skews the sample slightly towards larger firms. A reason for this might be that ESG rating agencies have better coverage for larger firms.

Table 3 shows the correlations between the ESG scores. For example, in the U.S., correlations range from -0.45 for the Refinitiv-RepRisk pair to 0.7 for the Moody's-ISS pair. The pair-wise correlations are fairly similar across the four currency regions.

RepRisk scores stand out for being correlated negatively with most other scores, suggesting that this rater employs a methodology that is markedly distinct. Indeed, a unique feature

Table 2. Descriptive Statistics. This table shows the descriptive statistics of all four subsamples: Eurozone, U.K., Japan, and the U.S. We use MSCI's IVA Industry Weighted score, Sustainalytics' ESG Risk Ratings, Refinitiv's TRESG score, RepRisk's Reputation Risk Index (RRI), Truvalue Labs' Insight Score, Moody's Global score, S&P Global's ESG score, and ISS's Numeric ESG Overall Rating. We multiplied Sustainalytics' and RepRisk' scores by -1 and added 100 so that a higher value corresponds to a better ESG performance for all ratings. *Return* is the monthly returns in percentage, *Beta* is the market beta estimated from monthly returns from month -60 to month -1, *Dividends* are the dividends per share over the prior 12 months divided by price at the end of the prior month, *Market Value* is the logarithm of the market value of equity at the end of the prior month, *Book-to-market* is the logarithm of book equity minus the logarithm of market value of equity at the end of the prior month, *Asset Growth* is the logarithm of growth in total assets in the prior fiscal year, *ROA* is the income before extraordinary items divided by average total assets in the prior fiscal year, *Momentum* is the return from month -12 to month -2, and *Volatility* is the monthly standard deviation, estimated from daily returns from month -12 to month -1. The sample consists of 273 firms and 15531 firm-month observations for the eurozone, 131 firms and 6655 firm-month observations for the U.K., 246 firms and 16147 firm-month observations for Japan, and 506 firms and 26974 firm-month observations for the U.S. All financial variables are winsorized at the 1% level. *Mean* corresponds to the mean and *StDev* to the standard deviation.

	Eurozone		U.K.		Japan		U.S.	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
ISS	2.18	0.40	1.98	0.38	1.73	0.33	1.73	0.36
Moody's	46.32	11.76	41.87	8.92	30.00	10.24	32.39	7.60
MSCI	6.71	2.06	6.90	1.82	5.39	1.99	4.82	2.17
Refinitiv	64.94	16.83	59.45	15.31	51.97	19.16	56.13	18.11
RepRisk	68.75	16.49	73.77	17.36	77.21	15.21	67.22	14.96
S&P Global	53.12	23.51	44.02	20.96	41.00	23.35	36.47	19.29
Sustainalytics	76.22	8.41	75.84	8.87	72.90	9.49	72.72	10.00
TVL	58.10	12.06	55.22	11.51	57.05	12.68	53.49	10.72
Return	0.87	8.37	0.67	9.01	0.58	8.42	0.94	8.64
Beta	0.91	0.39	0.88	0.40	0.96	0.36	1.07	0.51
Dividends	0.03	0.02	0.03	0.02	0.02	0.01	0.02	0.02
Market Value	9.12	1.14	8.58	1.13	13.71	0.93	9.81	1.20
Book-to-market	-0.76	0.72	-1.02	0.96	-0.38	0.60	-1.09	0.88
Asset Growth	0.05	0.14	0.08	0.18	0.05	0.09	0.06	0.16
ROA	0.04	0.05	0.06	0.07	0.04	0.04	0.06	0.06
Momentum	0.05	0.25	0.03	0.26	0.05	0.26	0.05	0.24
Volatility	0.06	0.02	0.07	0.03	0.06	0.02	0.07	0.03

Table 3. Correlations between ESG Scores. This table presents the correlations of all four subsamples: Eurozone, U.K., Japan, and the U.S. We use MSCI's IVA Industry Weighted score, Sustainalytics' ESG Risk Ratings, Refinitiv's TRESG score, RepRisk's Reputation Risk Index (RRI), Truvalue Labs' Insight Score (TVL), Moody's Global score, S&P Global's ESG score, and ISS's Numeric ESG Overall Rating. We multiplied Sustainalytics' and RepRisk' scores by -1 and added 100 so that a higher value corresponds to a better ESG performance for all ratings.

	ISS	Moody's	MSCI	Refinitiv	RepRisk	Sustainalytics	S&P Global	TVL
Eurozone								
ISS	1							
Moody's	0.65	1						
MSCI	0.47	0.50	1					
Refinitiv	0.56	0.63	0.42	1				
RepRisk	-0.24	-0.35	-0.07	-0.43	1			
S&P Global	0.50	0.59	0.39	0.63	-0.43	1		
Sustainalytics	0.26	0.31	0.39	0.23	0.18	0.22	1	
TVL	0.24	0.17	0.21	0.06	0.10	0.05	0.06	1
U.K.								
ISS	1							
Moody's	0.68	1						
MSCI	0.35	0.23	1					
Refinitiv	0.62	0.56	0.21	1				
RepRisk	-0.29	-0.31	0.05	-0.38	1			
S&P Global	0.55	0.64	0.18	0.68	-0.37	1		
Sustainalytics	0.23	0.18	0.39	0.10	0.17	0.16	1	
TVL	-0.02	-0.09	0.15	-0.04	0.30	-0.17	0.09	1
Japan								
ISS	1							
Moody's	0.61	1						
MSCI	0.38	0.41	1					
Refinitiv	0.56	0.69	0.34	1				
RepRisk	-0.21	-0.29	0.02	-0.36	1			
S&P Global	0.56	0.63	0.36	0.63	-0.34	1		
Sustainalytics	0.25	0.25	0.24	0.26	0.07	0.33	1	
TVL	0.07	0.11	0.10	0.10	0.11	0.06	0.05	1
U.S.								
ISS	1							
Moody's	0.70	1						
MSCI	0.40	0.38	1					
Refinitiv	0.63	0.68	0.36	1				
RepRisk	-0.33	-0.40	-0.10	-0.45	1			
S&P Global	0.59	0.63	0.31	0.64	-0.41	1		
Sustainalytics	0.14	0.08	0.23	0.18	0.12	0.10	1	
TVL	0.12	0.09	0.25	0.07	0.14	0.05	0.01	1

of the RepRisk score is that its purpose is to measure ESG risk with a focus on negative ESG incidents.[6] A likely explanation is that firms with lots of negative ESG headlines tend to invest heavily in their ESG reporting (Strike, Gao, and Bansal, 2006). Thus, firms with high ESG risk according to RepRisk may at the same time have good ratings from other providers that put more weight on firms' own reporting.

3 Errors-in-Variables and Stock Returns

In this section, we address the problem of noise in ESG ratings empirically. To do this, we examine the impact of the ESG measurement errors on the relationship between ESG performance and stock returns. We depart from the rapidly growing literature on the impact of ESG on financial outcomes in that we do not assume that ESG ratings provide accurate measurements. Instead, we assume that they measure ESG performance with noise.

A simplified asset pricing representation of stock returns and their relation to ESG performance can be written as follows:

$$r_{k,t+1} = \alpha + \beta \cdot Y_{k,t} + M_{k,t} + \epsilon_{k,t}, \tag{1}$$

where $r_{k,t+1}$ is the stock return between time t and $t+1$ for firm k. $Y_{k,t}$ is the true ESG performance at time t, $M_{k,t}$ is an omitted variable that affects stock returns and is correlated with ESG performance, and $\epsilon_{k,t}$ are the innovations assumed to be orthogonal to all the regressors. In Section 5, in which we estimate the effects of ESG performance on stock returns, we will additionally control for standard asset pricing characteristics.

The ESG performance $Y_{k,t}$ can be correlated with the omitted variable $M_{k,t}$. The omitted variable can be interpreted in many different ways: (i) as unexpected capital flows into ESG stocks, (ii) as investor preferences that shift toward ESG stocks, or (iii) as shifts in management quality. The presence of the omitted variable is important for interpreting the coefficients. We limit ourselves to addressing attenuation bias and design our empirical method to be robust to the presence of omitted variable bias.

In the previous section, we argued that ESG ratings are noisy. Therefore, in our specification below we assume that ESG rating agencies produce an imperfect measurement of the true ESG performance. Suppose that there are N ESG rating agencies indexed by i. The score of rating agency i for firm k is given by $s_{k,t,i}$ and this score contains measurement error

[6] For details, see https://www.reprisk.com/news-research/resources/methodology.

(or noise), denoted as $\eta_{k,t,i}$. Formally,

$$s_{k,t,i} = Y_{k,t} + \eta_{k,t,i}, \qquad i \in \{1, \ldots, N\}. \tag{2}$$

We assume the measurement error ($\eta_{k,t,i}$) is as in the classical errors-in-variables problem, that is, orthogonal to (Y, M, ϵ). (A complete list of assumptions follows in Section 3.1).

While the true ESG performance is not observable in the data, ESG scores are. The reduced-form of the structural model (1) that one can take to the data is as follows:

$$r_{k,t+1} = \alpha + \beta \cdot s_{k,t,i} + \nu_{k,t}, \tag{3}$$

where $\nu_{k,t} = M_{k,t} + \epsilon_{k,t} - \eta_{k,t,i} \cdot \beta$. In other words, we have replaced the true ESG performance Y with an ESG score, which is a noisy version of Y.

The coefficient of interest is β, whose OLS estimate is given by

$$\beta_{OLS} = \frac{var(Y_{k,t})}{var(Y_{k,t}) + var(\eta_{k,t,i})} \cdot \left[\beta + \frac{cov(Y_{k,t}, M_{k,t})}{var(Y_{k,t})}\right] \tag{4}$$

It is easy to see that the OLS estimate of β is biased for two reasons. The first bias is the attenuation bias, $var(Y_{k,t})/(var(Y_{k,t}) + var(\eta_{k,t,i}))$, which occurs because of the measurement error in the regressor, while the second bias is the omitted variable bias, $cov(Y_{k,t}, M_{k,t})/var(Y_{k,t})$.

We concentrate on the attenuation bias as opposed to the omitted variable bias for two reasons. First, given the substantial disagreement of ESG scores across rating agencies, noise in ESG scores is a first-order problem for regulators, asset managers, and investors in general. Second, the fact that there are multiple ESG ratings that disagree, but are nonetheless correlated with each other, provides a unique opportunity to address the attenuation bias that results from noisy measurement.

One of the known approaches for tackling attenuation bias is to use alternative noisy measures of the regressor as instruments. Consider again the reduced-form regression (3), in which the score $s_{k,t,i}$ of ESG rating agency i is a noisy measure of Y_k, t. We will use the scores of other ESG ratings as instruments. In the empirical section, we will use several instruments, but here, for expositional purposes, let us focus on one instrument $s_{k,t,j}$, which is the score of firm k at time t from rater j (for $j \neq i$).

For the moment, assume that $s_{k,t,j}$ is a valid instrument for $s_{k,t,i}$. We will discuss this assumption in detail in Section 3.1 below. The IV estimate of β is then given by

$$\beta_{IV} = \left[\beta + \frac{cov(Y_{k,t}, M_{k,t})}{var(Y_{k,t})}\right] \tag{5}$$

It is instructive to compare the OLS and IV estimates (Equations (4) and (5)). Notice that the omitted variable biases both the OLS estimate and the IV estimate in exactly the same manner. Hence, to isolate the attenuation bias we simply need to compute the ratio of the two estimates. An implicit assumption behind the IV estimation is that the omitted variable bias does not affect the measurement error (we formalize this in the next section in assumption (8)). The magnitude of the noise in the ESG score, which is used as a regressor, can then be estimated as

$$\kappa_i \equiv 1 - \frac{\beta_{OLS}}{\beta_{IV}} = 1 - \frac{var(Y)}{var(Y) + var(\eta_i)} = \frac{var(\eta_i)}{var(Y) + var(\eta_i)}, \tag{6}$$

where κ_i is the noise-to-signal ratio in the rater i's scores. In Section 5.1, we will compute the noise-to-signal ratios and compare them across different raters.

3.1 Identifying assumptions for the IV procedure

Under which conditions are ESG ratings valid instruments for each other? In a nutshell, we need to assume that ESG ratings are related to each other *only* through $Y_{k,t}$ and that their measurement error is white noise. We now formally spell out the assumptions under which ESG ratings are valid instruments for each other. We do so for the case in which there are multiple instruments.

As a baseline, we need the *relevance* assumption, i.e., that instruments are correlated with the regressor. This assumption is easy to defend. In Section 2, we demonstrate that ESG rating scores of different agencies are positively correlated and that some of these correlations are relatively high. In Section 5, we will show formally that we do not have a problem of weak instruments (see the first-stage F-statistics in Tables 4–6).

Given relevance, we require three further assumptions regarding the measurement errors $\eta_{k,t,i}$. First, the errors are *classical*, i.e., additive and orthogonal to Y, as in the classical errors-in-variables problem:

$$E[\eta_{k,t,i} | Y_{k,t}] = 0, \quad \forall i. \tag{7}$$

Second, the error terms $\eta_{k,t,i}$ are independent of both the stock cash-flow innovations $\epsilon_{k,t}$ (not captured by the firm-level controls that we introduce later) and the omitted variable $M_{k,t}$, i.e.,

$$E[\eta_{k,t,i} \cdot \epsilon_{k,t}] = 0 \quad \text{and} \quad E[\eta_{k,t,i} \cdot M_{k,t}] = 0, \quad \forall i. \tag{8}$$

Furthermore, when we introduce controls $X_{k,t}$ in our baseline regression, we will assume that $E[\eta_{k,t,i} \cdot X_{k,t}] = 0$. These three assumptions are the *exclusion restriction.*

Third, we assume that all errors $(\eta_{k,t,i})$ are independent across rating agencies:

$$E[\eta_{k,t,i} \cdot \eta_{k,t,j}] = 0 \quad \forall j \neq i. \tag{9}$$

This is the *independence* assumption.

Assumptions (7), (8), and (9) imply that the measurement error of each ESG rating is effectively white noise. This is a strong assumption, and we review below some possibilities regarding how it could be violated. While we cannot rule out those violations in principle, we can test for them empirically using the Sargan-Hansen OIR test, as we explain in Section 3.2.

The most probable threat to our IV estimation is a violation of the independence assumption (9), i.e., noise in ESG scores may be correlated across raters. This can occur if several rating agencies use similar data and similar estimation procedures to arrive at their scores. Also, rating agencies may rely on the same imputation method for missing data. As imputation always approximates the missing true value with some error, this error would then be correlated across those ratings that use the same procedure. It is also possible, although less likely, that ESG rating agencies retroactively adapt their scores after observing contemporaneous and past stock return realizations. This would result in a violation of the exclusion restriction (8). Finally, errors could be non-classical, causing a violation of assumption (7). This could be the case if errors are related to the true ESG performance in an asymmetric or non-linear fashion. For instance, it could be that bad performance is easier to detect than good performance.

It is also possible that the errors are correlated with the omitted variable, thus violating assumption (8). For instance, managers that foster a collaborative work environment might also score high on diversity through inclusiveness. This could affect both, the ESG rating and the returns of the firm. This violation would not be diagnosed by the OIR test. In this case, our procedure will still tackle the attenuation bias, even though the interpretation of the coefficient changes. As argued before (using Equations (4) and (5)), even though the IV estimate may suffer from the omitted variable bias, the ratio between the OLS and the IV estimates remains unaffected.

3.2 Testing the Validity of Instruments

In our setting we have several ESG rating agencies producing ratings that intend to capture a firm's true ESG performance. This implies that there are several rating agencies that could be used as instruments. When there are two or more instruments, it becomes possible to run overidentifying restriction tests and thereby check the validity of the identifying assumptions.

The first step is to argue that different ESG ratings can be used as both regressors and instruments. This is a counter-intuitive implication of the errors-in-variables setting that is sometimes missed in applied work. In many applications, the regressors and the instruments are not interchangeable. In the case of errors-in-variables, they are, as long as the instruments are valid. This is because model misspecification is not in the structural form but in the measurement of the variables.[7]

Let us now show what happens to the OLS and IV estimates if assumptions (7), (8), and (9) are violated. Suppose that the measurement error in the regressor s_i is correlated with the true ESG performance, with the omitted variable, with the stock market innovations, and with the measurement errors of another rating agency s_j, which we use as an instrument. The OLS and the IV estimates are given by (we have suppressed the notation $\{k, t\}$ for simplicity):

$$\beta_{OLS} = \frac{\beta \cdot var(Y) + cov(Y, M) + \beta \cdot cov(Y, \eta_i) + cov(M, \eta_i) + cov(\epsilon, \eta_i)}{var(Y) + var(\eta_i) + cov(Y, \eta_i)},$$

$$\beta_{IV} = \frac{\beta \cdot var(Y) + cov(Y, M) + \beta \cdot cov(Y, \eta_j) + cov(M, \eta_j) + cov(\epsilon, \eta_j)}{var(Y) + cov(Y, \eta_i) + cov(Y, \eta_j) + cov(\eta_i, \eta_j)},$$

where the term in purple is what causes attenuation bias in the OLS estimate, the terms in red are the ones due to violation of the classical errors-in-variables (Equation (7)), the term in green is due to the violation of the exclusion restriction (Equation (8)), and the term in blue corresponds to the violation of the independence of the instruments (Equation (9)).[8]

The OIR test compares IV estimates from two models, with different sets of instruments. If the identifying assumptions (7), (8), and (9) hold, the two IV estimates should be the same; otherwise, they are different. Denote the ESG score used in the regression (3) as i

[7] A simple example that highlights the intuition is as follows. Assume that we are interested in estimating the impact of a variable x on $y = \beta x + \epsilon$, where the error and the regressor are orthogonal. If we can observe x perfectly, the OLS estimate is consistent and unbiased. Now assume that we observe not the variable x but two noisy versions of it, $s_1 = x + \eta_1$ and $s_2 = x + \eta_2$, and the errors η_1 and η_2 are white noise. The bias comes from the substitution of s_1 for x in the regression equation: $x = s_1 - \eta_1$ and $y = \beta s_1 + \epsilon - \beta \eta_1$. Because s_1 and η_1 are correlated, now the OLS estimate is inconsistent and biased. If the measurement errors are orthogonal to each other, s_2 is correlated with s_1 through x, but, under our assumptions, their errors are uncorrelated. Therefore, we can use s_2 as an instrument of s_1 and vice versa. The identifying assumptions that the measurement errors are uncorrelated with all innovations and x are true for both variables.

[8] See Appendix A.1 for the derivations.

and the two instruments as s_{j_1} and s_{j_2}:

$$s_{j_1} = Y + \eta_{j_1},$$
$$s_{j_2} = Y + \eta_{j_2}.$$

The IV estimators from the two models that use s_{j_1} and s_{j_2}, respectively, as instruments are given by

$$\begin{aligned}
\beta_{IV_1} &= \frac{\beta \cdot var(Y) + cov(Y, M) + \beta \cdot cov(Y, \eta_{j_1}) + cov(M, \eta_{j_1}) + cov(\epsilon, \eta_{j_1})}{var(Y) + cov(Y, \eta_i) + cov(Y, \eta_{j_1}) + cov(\eta_i, \eta_{j_1})} \\
\beta_{IV_2} &= \frac{\beta \cdot var(Y) + cov(Y, M) + \beta \cdot cov(Y, \eta_{j_2}) + cov(M, \eta_{j_2}) + cov(\epsilon, \eta_{j_2})}{var(Y) + cov(Y, \eta_i) + cov(Y, \eta_{j_2}) + cov(\eta_i, \eta_{j_2})}
\end{aligned} \tag{10}$$

Notice that under the identifying assumptions (7), (8), and (9) the two IV estimates are identical,

$$\beta_{IV_1} = \beta_{IV_2} = \beta + \frac{cov(Y, M)}{var(Y)},$$

and equal to the estimate in Equation (5). However, when the identifying assumptions fail, the two IV estimates are different from each other:[9]

$$\beta_{IV_1} \neq \beta_{IV_2}$$

The OIR test tests the equality of the two IV estimates.

We implement this test using the Sargan-Hansen OIR test. The Sargan-Hansen test uses the two instruments (or all the available instruments) simultaneously in the first-stage regression and then compares the correlations between the instruments and the residuals from the regression.

We discussed at the end of Section 3.1 possible economic reasons why our identifying assumptions could be violated. Unfortunately, the OIR test cannot diagnose which assumption is violated, i.e., it cannot determine which of the covariances is different from zero; instead we can speculate about this based on economic arguments. However, the OIR test indicates which instruments cause violations, thus allowing us to estimate the model with a subset of instruments for which the OIR test does not reject the model.

[9]There is one knife-edge possibility that all the covariances in Equations (10) are different from zero but identical across the two models, in which the the two IV estimates are identical.

3.3 Estimation Procedures

In our empirical implementation we find that of the many instruments we have available, some are valid and some are not. How do we choose, then, which instruments to include in the estimation? The Sargan-Hansen OIR test unfortunately does not discern which instrument is valid—in essence, it finds that the coefficients are different, but it does not identify which coefficient is right.

We propose two procedures. The first procedure, which we term *Pruning*, starts with the full set of instruments and reduces it as the OIR test is rejected. The second procedure, called *Lasso*, starts with a minimal set of instruments and adds instruments as the OIR is not being rejected.

In the *Pruning* procedure, we select a rating agency whose scores we would like to instrument, i.e., the regressor. We use all remaining rating agencies' scores as instruments, so the IV estimator we have been discussing technically becomes a 2SLS estimator. Second, we estimate specification (3) using 2SLS and run the Sargan-Hansen OIR test. If the model passes the Sargan-Hansen test, then all included instruments are valid; otherwise, we exclude instruments, one at a time, until the model passes the test. We report the included and excluded instruments. The Pruning procedure identifies the maximum number of valid instruments for each ESG rating i, $i = 1, \ldots, 8$, and provides the valid 2SLS estimate of the effect of ESG performance i on stock returns.

There are two concerns, however, regarding the Pruning procedure. First, it is unclear whether the OIR tests have sufficient power in our application. The short answer is that they do. We indeed find many rejections in our empirical implementation (Section 5). Additionally, we evaluate the power of the OIR test in simulations (see Section 6) and confirm that the test has sufficient power to detect invalid instruments. Second, the Pruning procedure is a sequence of OIR tests and the size should be adjusted to reflect the fact that the tests are not independent. We indeed estimate many OIR tests in this procedure. For example, when 7 ESG ratings are included as instruments, we run only one OIR test. If the OIR test rejects that model, our next step is to perform the OIR tests on 7 combinations of 6 instruments. If all of these 7 tests reject the model, we consider 21 possible combinations of 5 instruments and therefore run 21 OIR tests and so forth until a set of instruments passes the test. To address this, we select a very strict rejection threshold of 1 percent in two-sided tests. We have estimated even tighter bands and the results are virtually identical. This gives us confidence that the model that passes the OIR tests satisfies our identifying assumptions (7), (8), and (9).

Our second procedure, which we call *Lasso*, takes a different approach, one that reduces the problem of repeated tests. In the first stage, it estimates a Lasso regression with a very large penalty — so large that only one rating agency's set of scores is chosen as an instrument. We then reduce the penalty until the set of scores of a second rating agency are chosen as another instrument. Since we now have more than one instrument, we can run the OIR test. If the model passes it, we continue decreasing the penalty and adding the scores of one rating agency at a time. Notice that in this procedure there is a maximum of 7 OIR tests.

We expect the coefficients to be similar between the Pruning and the Lasso procedures, but, because of the attenuation bias, quite different from the OLS. We show in Section 5 that this is indeed the case.

3.4 Different Horizons

Following Pancost and Schaller (2021), we evaluate the robustness of the errors-in-variables estimates by changing the dependent variable.[10] In our application, if the stock returns are computed at different horizons (1 month, 2 months, etc.), one can argue that the variances of the innovations change, that the importance of the omitted variable shifts, etc. One feature that is common to all these specifications, however, is the magnitude of the attenuation bias, as we show below.

Denote the (monthly) stock return from time t until $t+h$ by $r_{k,t+h}$, where h is the horizon over which the return is measured. The structural forms of the relationship between stock returns $r_{k,t+h}$, for each horizon h, and ESG performance are

$$
\begin{aligned}
r_{k,t+1} &= \alpha_1 + \beta_1 \cdot Y_{k,t} + M_{1,k,t} + \epsilon_{1,k,t}, \\
r_{k,t+2} &= \alpha_2 + \beta_2 \cdot Y_{k,t} + M_{2,k,t} + \epsilon_{2,k,t}, \\
&\vdots \\
r_{k,t+h} &= \alpha_h + \beta_h \cdot Y_{k,t} + M_{h,k,t} + \epsilon_{h,k,t}.
\end{aligned}
$$

Even though we use the same ESG performance for each of the different return horizons, we allow for the omitted variable to change with the horizon. The reduced forms of the above

[10] We thank Aaron Pancost for suggesting this test to us.

structural equations are

$$
\begin{aligned}
r_{k,t+1} =& \alpha_1 + \beta_1 \cdot s_{k,t,i} + \nu_{1,k,t}, \\
r_{k,t+2} =& \alpha_2 + \beta_2 \cdot s_{k,t,i} + \nu_{2,k,t}, \\
&\vdots \\
r_{k,t+h} =& \alpha_h + \beta_h \cdot s_{k,t,i} + \nu_{h,k,t}.
\end{aligned}
$$

The regressors are the same across specifications, and we allow the coefficients, the errors, and the omitted variable to have different variances and covariances. This implies that the covariance between the ESG score and the omitted variable is also likely to change.

The OLS and the IV estimates for any horizon h are given by

$$
\beta_{OLS,h} = \left[\frac{var(Y_{k,t})}{var(Y_{k,t}) + var(\eta_{k,t,i})}\right] \left[\beta_h + \frac{cov(Y_{k,t}, M_{h,k,t})}{var(Y_{k,t})}\right], \tag{11}
$$

$$
\beta_{IV,h} = \left[\beta_h + \frac{cov(Y_{k,t}, M_{h,k,t})}{var(Y_{k,t})}\right]. \tag{12}
$$

It is entirely possible that $\beta_{OLS,1} \neq \beta_{OLS,h}$ and that $\beta_{IV,1} \neq \beta_{IV,h}$. However, the ratios of the IV and the OLS coefficients (which are our estimates of the attenuation bias) are identical across all horizons.

$$
\frac{\beta_{OLS,1}}{\beta_{IV,1}} = \frac{\beta_{OLS,2}}{\beta_{IV,2}} = \frac{\beta_{OLS,h}}{\beta_{IV,h}} = \left[\frac{var(Y_{k,t})}{var(Y_{k,t}) + var(\eta_{k,t,i})}\right]. \tag{13}
$$

Of course, if there is misspecification in the instruments, then the ratio will not remain the same across the entire range of our left-hand-side variables. In the empirical implementation (Section 5), we change the horizon over which the stock returns are measured and evaluate how stable the ratio of the IV and the OLS coefficients is.

4 Model

In this section, we present a simple, stylized model with traditional and ESG investors that highlights the attenuation effect when ESG signals are noisy. We abstract from the omitted variable problem. Our focus is on a single ESG signal, measured with noise. We will show that such measurement error leads to bias in a standard regression analysis of the relationship between stock returns and ESG performance. The noisier the ESG signal, the larger the bias.

We consider a two-period model, with $t = 0, 1$. Investment opportunities are represented by a risky stock of a single firm and a riskless bond, with the risk-free rate normalized to zero.[11] The stock is a claim to the cash flow $D \sim N(\overline{D}, \sigma_D^2)$ per share, with D realized in period 1. The stock is in fixed supply of $\overline{\theta}$ shares and the riskless bond is in infinite net supply. We denote the stock price in period t by p_t, where $p_1 = D$.

There is a measure λ of ESG investors and $1 - \lambda$ of traditional investors. Both types of agents invest their funds into the stock and the bond. The ESG and traditional investors' portfolio allocation to the stock is θ^i, where $i = ESG, T$, respectively. The period-1 wealth of the investors W_1^i is then $W_0^i + \theta^i(D - p_0)$, where W_0^i is their initial wealth, $i = ESG, T$. ESG investors derive a non-pecuniary benefit Y per share from holding the stock, with $Y \sim N(\overline{Y}, \sigma_Y^2)$ independent from D. Their utility is exponential, $U(W_1, Y) = -exp(-\gamma(W_1 + \theta^{ESG} Y))$.[12] We think of Y as an ESG externality, generated by the firm, which ESG investors internalize. The traditional investors have utility $U(W_1) = -exp(-\gamma W_1)$ and do not internalize any ESG externalities. Investors' initial endowments are in terms of shares of the stock and bond and they choose their portfolios to maximize their expected utilities.

In period 0, investors receive noisy signals, s_D and s_Y, about cash flows and ESG benefit, D and Y, respectively:

$$s_D = D + \eta_D, \tag{14}$$
$$s_Y = Y + \eta_Y, \tag{15}$$

where $\eta_i \sim N(0, \sigma_{\eta_i}^2),\ i = D, Y$ are independent of each other and independent of D and Y.

4.1 Portfolio Choice and Asset Prices

To solve for equilibrium, we first need to solve the inference problem of the investors. Exploiting the joint normality of random variables in our economy, we arrive at the following lemma (all proofs cam be found in the Appendix A.2).

Lemma 1 *The mean and variance of D, conditional on signal s_D, are given by*

$$E(D|s_D) = \overline{D} + \beta(s_D - \overline{D}) = \overline{D} + \frac{\sigma_D^2}{\sigma_D^2 + \sigma_{\eta_D}^2}(s_D - \overline{D}), \tag{16}$$
$$Var(D|s_D) = \sigma_{\nu_D}^2 = \frac{\sigma_D^2 \sigma_{\eta_D}^2}{\sigma_D^2 + \sigma_{\eta_D}^2}. \tag{17}$$

[11] It is straightforward to extend the model to multiple risky stocks.

[12] Our approach to modeling ESG investors is similar to that of Pastor, Stambaugh, and Taylor (2021b) and Friedman and Heinle (2016).

The mean and variance of Y, conditional on signal s_Y, are as follows:

$$E(Y|s_Y) = \overline{Y} + \beta(s_Y - \overline{Y}) = \overline{Y} + \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{\eta Y}^2}(s_Y - \overline{Y}), \tag{18}$$

$$Var(Y|s_Y) = \sigma_{\eta Y}^2 = \frac{\sigma_Y^2 \sigma_{\eta Y}^2}{\sigma_Y^2 + \sigma_{\eta Y}^2}. \tag{19}$$

We are now able to solve for optimal portfolios of ESG and traditional investors. These portfolios are given by

Lemma 2 (Portfolio Choice) *The investors portfolio demands are*

$$\theta^T = \frac{1}{\gamma}\frac{E(D|s_D) - p_0}{Var(D|s_D)}, \tag{20}$$

$$\theta^{ESG} = \frac{1}{\gamma}\frac{E(D|s_D) + E(Y|s_Y) - p_0}{Var(D|s_D) + Var(Y|s_Y)}. \tag{21}$$

The traditional investors hold the standard mean-variance portfolio, which optimally trades off risk (the denominator) and expected return (the numerator). In contrast, ESG investors account for ESG characteristics in their portfolio choice. The higher the stock's expected ESG benefit Y, the more shares of it ESG investors are willing to include in their portfolio. However, since ESG investors are risk-averse, the perceived risk of the stock is higher for them relative to traditional investors. This additional risk is driven by the noise in ESG ratings—the higher this noise, the less of the stock ESG investors are willing to hold (see the denominator of the portfolio demand in (21)).

The market clearing condition requires that investors' demand for the stock equals its supply, i.e.,

$$\lambda\theta^{ESG} + (1 - \lambda)\theta^T = \overline{\theta}. \tag{22}$$

To solve for the equilibrium stock price, we substitute the optimal portfolios from Lemma 2 into the market clearing condition (22). We report the resulting period-0 stock price in the following proposition.

Proposition 1 (Asset Prices) *The period-0 stock price is given by*

$$p_0 = \overline{D} + \frac{\sigma_D^2}{\sigma_D^2 + \sigma_{\eta D}^2}(s_D - \overline{D}) + A\lambda\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2}\left[\overline{Y} + \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{\eta Y}^2}(s_Y - \overline{Y})\right] \tag{23}$$

$$- A\gamma\overline{\theta}\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2}\left[\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} + \frac{\sigma_Y^2\sigma_{\eta Y}^2}{\sigma_Y^2 + \sigma_{\eta Y}^2}\right], \tag{24}$$

where $A = \left[\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2} + (1-\lambda)\frac{\sigma_Y^2\sigma_{\eta Y}^2}{\sigma_Y^2+\sigma_{\eta Y}^2}\right]^{-1}$.

The ESG performance Y does not affect fundamentals (i.e., the firm's cash flow D). However, it does affect asset prices because there is a group of investors that care about it. A positive signal s_Y about the ESG performance Y boosts the stock price. Y can be interpreted as the true ESG performance, and s_Y as what the ESG rating agencies measure — their scores.

Suppose that the stock is a green stock, which appeals to ESG investors, i.e., $\overline{Y}$ is positive and sufficiently high. Then, relative to an economy with no ESG investors, the stock price will be higher, reflecting the additional benefit to ESG investors from holding a green stock. The mass of ESG investors λ is another important parameter. The higher the mass of ESG investors, the higher the stock price.

Let us now examine a *realized* per-share return on the stock in period 0:

$$\begin{aligned} p_0 - p_{-1} =& \overline{D} - S_{-1} + \frac{\sigma_D^2}{\sigma_D^2+\sigma_{\eta D}^2}(s_D - \overline{D}) + A\lambda\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2}\left[\overline{Y} + \underbrace{\frac{\sigma_Y^2}{\sigma_Y^2+\sigma_{\eta Y}^2}}_{\text{attenuation effect}}(s_Y - \overline{Y})\right] \\ & - A\gamma\overline{\theta}\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2}\left[\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2} + \frac{\sigma_Y^2\sigma_{\eta Y}^2}{\sigma_Y^2+\sigma_{\eta Y}^2}\right]. \end{aligned} \tag{25}$$

We think about the constant p_{-1} as the value of the stock one period before ESG investors (unexpectedly) arrived in the market. The realized returns on the stock depends on the magnitude of (unanticipated) ESG investor inflows, captured by λ, with the inflows boosting returns of green stocks. In contrast, stock returns of brown firms (firms with a sufficiently negative ESG benefit Y) fall.

We now present the expression for the *expected* per-share return on the stock.

$$\begin{aligned} E(D) - p_0 =& -\frac{\sigma_D^2}{\sigma_D^2+\sigma_{\eta D}^2}(s_D - \overline{D}) - A\lambda\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2}\left[\overline{Y} + \underbrace{\frac{\sigma_Y^2}{\sigma_Y^2+\sigma_{\eta Y}^2}}_{\text{attenuation effect}}(s_Y - \overline{Y})\right] \\ & + A\gamma\overline{\theta}\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2}\left[\frac{\sigma_D^2\sigma_{\eta D}^2}{\sigma_D^2+\sigma_{\eta D}^2} + \frac{\sigma_Y^2\sigma_{\eta Y}^2}{\sigma_Y^2+\sigma_{\eta Y}^2}\right]. \end{aligned} \tag{26}$$

The higher the ESG signal s_Y, the lower the expected return on the stock. This is because in our model the firm's cash flow D is fixed and therefore the more ESG investors push up the stock price in response to a high ESG signal, the lower the stock's expected return going forward. That is, the effects of ESG on stock prices in our model manifest themselves entirely through the cost of capital channel.[13]

Both the expected and realized returns depend on the noise in the ESG signal, σ_{η_Y}. The noisier the signal s_Y, the lower its passthrough to stock returns. Put differently, noise in the signal creates an attenuation effect (see the highlighted term in Equations (25) and (26)). In the limit of $\sigma_{\eta_Y} \to \infty$, the effect of s_Y on stock returns is fully attenuated. These observations will become important in our empirical analysis, which uses data on ESG scores, which we interpret as noisy ESG signals. The noise in the reported ESG scores is apparent from the discrepancies in measuring the same ESG performance by different ratings providers. In the next section, we treat this as a classical errors-in-variables problem and propose a procedure that tackles the attenuation bias in standard regressions of stock returns on noisy measures of ESG performance.

5 Empirical Results

In this section, we estimate the OLS regressions of stock returns on ESG ratings and contrast them to 2SLS regressions, which use scores of other rating agencies as instruments. We discuss the results for the 1-, 2-, and 3-month returns, and provide additional results for 4- to 8- month returns in the appendix.

The baseline regression is

$$r_{k,t+h} = \alpha + \beta \cdot s_{k,t,i} + c_X \cdot X_{k,t} + \nu_{h,k,t}, \tag{27}$$

where $s_{k,t,i}$ denotes the ESG rating of firm k, by rater i, in month t, h denotes the horizon (in months), and all returns are monthly. Following Lewellen (2015), we include stock-level controls $X_{k,t}$ consisting of *Beta*, *Dividends*, *Market Value*, *Book-to-market*, *Asset Growth*, *ROA*, *Momentum*, and *Volatility*. $X_{k,t}$ also includes industry and month fixed effects.[14] We cluster standard errors by month and the GICS sub-industry.

As argued in Sections 3 and 4, the OLS estimate of the effect of ESG performance on stock returns, β_{OLS}, suffers from attenuation bias. To assess the significance of the bias, we compare the OLS estimates with their 2SLS counterparts. The first-stage regression uses

[13] We abstract away from the cash flow risk channel by assuming that the stock's future cash flow D is uncorrelated with the ESG characteristic Y. However, it is entirely possible that firms with low ESG performance are riskier than their greener counterparts (e.g., regulation risk) and therefore their expected returns are higher.

[14] We do not use firm fixed effects because the frequency of most ESG ratings changes is annual, and we have a very limited time series.

ESG scores of other rating agencies as instruments for a given rater's score and includes the same controls as in (27):

$$s_{k,t,i} = c_0 + \pi \cdot Z_{k,t,i} + c_1 \cdot X_{k,t} + \eta_{k,t}, \tag{28}$$

where $Z_{k,t,i} := \{s_j | \forall (j \neq i),\ s_j \text{ is a valid instrument}\}$ are other rating agencies' scores that are being used as instruments. We determine the set of valid instruments by either the Pruning or Lasso procedures. Denote by $\hat{s}_{k,t,i}$ the fitted value from the estimation of Equation (28). Then the second stage regression is

$$r_{k,t+h} = \alpha + \beta \cdot \hat{s}_{k,t,i} + c_X \cdot X_{k,t} + \nu_{h,k,t}. \tag{29}$$

Provided that our assumptions are satisfied, we expect that $|\beta_{2SLS}| > |\beta_{OLS}|$. In the empirical implementation, we partial out all the controls from returns and the ESG rating scores, thereby, allowing us to use equations from Section 3 directly.

Table 4 reports the results for the eurozone, the UK, Japan, and the US. We first discuss OLS coefficients. Our estimation in Table 4 reveals that 9 of 32 OLS coefficients are significant (at the 10, 5, or 1 per cent level). Most of the coefficients are positive, although in the U.K. six are negative, and in the U.S. one is negative. Positive coefficients are not what one would expect based on the equilibrium asset pricing model presented in Section 4, but it is consistent with empirical findings covering a similar time period. For instance, Pastor, Stambaugh, and Taylor (2021a) also find positive returns for high ESG stocks attributing this to changes in climate concerns over the sample period, as well as unexpected flows into ESG stocks. Our methodology allows for an omitted variable in the regression; hence, we can estimate the attenuation bias even when capital flows are not included as a regressor. In unreported robustness checks, we have estimated the regression with annual flows, and results were virtually identical; in particular, the signs and significance of the coefficients were the same.

The second and most important finding is that the OLS estimator suffers from the attenuation bias. Using the Pruning IV procedure, 23 of 32 2SLS coefficients are larger (in absolute terms) than their OLS counterparts, and by a substantial amount. There is one case in which the coefficient switches sign (Moody's in the US). Excluding the sign-switching case, the average ratio between the 2SLS and the OLS coefficients is 4.77 and the median ratio is 2.08. That is, more than half of the coefficients double in size. Furthermore, 24 of the 2SLS coefficients are significant at the 10 or 5 per cent level. In many cases, the OLS coefficients do not only become larger in magnitude, but become significant after we reduce noise by applying the Pruning procedure. For example, the coefficient for Truvalue Labs is significant only after instrumentation in all four regions.

The Lasso procedure confirms the results of the Pruning procedure and delivers results consistent with the attenuation bias. 23 out of 32 coefficients are larger than their OLS counterparts. Excluding the three cases of sign-switching, the average ratio of 2SLS to OLS coefficients is 4.84 and the median ratio is 2.28. 25 coefficients are significant, and in most cases this corresponds to an increase of the coefficient. Thus, for both Pruned and Lasso procedures, there is a clear tendency for 2SLS coefficients to be larger than the corresponding OLS estimates.

The third important observation from Table 4 is that there are numerous instances in which one or more than one instrument is rejected by the Sargan-Hansen OIR test. Table 4 shows accepted instruments as green check marks and excluded instruments as red cross-outs. The pattern of rejections is different for Pruned IV and Lasso IV, and coefficients are sensitive to the selection of instruments. In the first such case, S&P Global in the eurozone, the OLS coefficient is 0.122, Pruned IV is 0.055, and Lasso IV is 0.103. However, none of the 2SLS coefficients is significant in this case. The bad news is that violations of the assumptions outlined in Section 3.1 exist, while the good news is that they are detected by the OIR test. In Section 6, we will provide evidence from simulations that show that, in our application, the OIR test is quite sensitive to violations of our identifying assumptions (7), (8), and (9). While the OIR test does not provide a diagnosis of the reasons for rejection, based on our discussion in Section 3.1, we believe that the most probable reason is that measurement errors are correlated across rating agencies.

The fourth observation is that, despite their low correlations with each other, ESG scores of other raters are strong instruments for a given ESG score. F-statistics in Table 4 range from 125 to 4300. This demonstrates strong support for our relevance assumption. While some individual scores could be very noisy and lack coverage of some ESG attributes, taken together, they work extremely well in predicting any given rater's scores. It is true, as Table 4 shows, that some instruments are invalid and therefore have to be dropped from the estimation procedure. However, there are enough valid instruments left to guarantee a strong first stage.

The results illustrate that in our 2SLS approach even a very noisy measure adds value. S&P Global and Truvalue Labs are two examples that do not produce significant OLS coefficients in Tables 4 to 6, yet are very rarely rejected as instruments. At the same time, after instrumenting with other ratings, the 2SLS coefficients of these two raters increase substantially and are often significant. This may indicate that these two ratings contain a substantial amount of noise, and yet convey essential information. Importantly, the noise appears to be orthogonal to the noise that is contained in other ratings. This pattern could well be related to these ratings' unique methodologies. Truvalue Labs employs a methodology that relies strongly on computer algorithms to process large amounts of online information. S&P Global's methodology employs a very detailed questionnaire collecting information directly from companies. Although it seems that these methodological features add noise, they also add important information when used in a 2SLS estimation.

Our analysis in Section 3.4 predicts that the attenuation bias should manifest itself regardless of the horizon at which stock returns are measured. The OLS and the 2SLS estimated coefficients may vary with the horizon, yet their ratio should remain stable. To test this prediction in the data, we repeat the regressions in Table 4 for 2- to 8-month stock returns as dependent variables.

Tables 5 and 6 present the estimation results for 2-month and 3-month stock returns as dependent variables, respectively. The tables reveal that the results are qualitatively the same and in some aspects are even stronger. For 2-month returns, Pruned IV yields 24 coefficients that go up compared to the corresponding OLS coefficients; all 24 of those are significant; the coefficients increase on average by a factor of 11.45 after instrumentation (strongly driven by Sustainalytics in the UK) and the median increase is a factor of 2.9. For 3-month returns, Pruned IV yields 26 coefficients that go up compared to OLS, 23 of which are significant. The results for Lasso IV are not quite as consistent, and in some cases no instrument passes the Sargan-Hansen OIR test, preventing us from estimating the relevant coefficient. Nonetheless, the majority of estimated coefficients increase compared to OLS. There are a few cases in which Pruned IV and Lasso IV would lead to different conclusions, but these cases are rare.[15] Overall, these results provide evidence that attenuation bias is present and that it is stable across alternative left-hand-side variables.

[15] There are two cases in which coefficients are significant but point in opposite directions. For Refinitiv, the OLS estimate based on 2-month returns in Japan is 0.187, Pruned IV is larger (0.220), but Lasso is smaller (0.150). For MSCI, the OLS estimate based on 3-month returns in the eurozone is 0.145, Pruned IV is larger (0.339), and Lasso is smaller (0.117).

Table 4. 1-month stock returns and ESG ratings. This table reports estimates of β from the OLS regression (27) and the 2SLS regression (29). The first set of columns shows OLS estimates, while the second and third set of columns show 2SLS coefficients from the Pruning and Lasso procedures, resp., described in Section 3.3. All returns are monthly and all reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. The checkmarks in the columns titled "Valid IV" indicate a selection of instruments that passes the Sargan-Hansen OIR test. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

Region	Rater	OLS			Pruned IV													Lasso IV											
		Coeffs	StdErr		Coeffs	StdErr		Valid IV								Ftest	Coeffs	StdErr		Valid IV								Ftest
EUR	ISS	0.152	0.071	**	0.137	0.054	***		✓	✓	✓	✓	✓	✓	✓	1258	0.137	0.054	***		✓	✓	✓	✓	✓	✓	✓	1258
	MSCI	0.123	0.053	**	0.194	0.071	***	✓		✓	✓	✓	✓	✓	✓	1029	0.194	0.071	***	✓		✓	✓	✓	✓	✓	✓	1029
	Refinitiv	0.012	0.066		0.183	0.071	***	✓	✓		✓	✓	✓	✓	✓	1660	0.183	0.071	***	✓	✓		✓	✓	✓	✓	✓	1660
	RepRisk	0.146	0.069	**	0.333	0.072	***	✓	✓	✓		✓	✓	✓	✓	432	0.333	0.072	***	✓	✓	✓		✓	✓	✓	✓	432
	S&P Global	0.122	0.083		0.055	0.055		✗	✓	✓	✓		✓	✓	✓	1296	0.103	0.072		✓	✓	✓	✓		✓	✓	✓	1178
	Sustainalytics	0.123	0.076		0.237	0.071	***	✓	✓	✓	✓	✓		✓	✓	881	0.237	0.071	***	✓	✓	✓	✓	✓		✓	✓	881
	TVL	0.039	0.066		0.634	0.071	***	✓	✓	✓	✓	✓	✓		✓	125	0.634	0.071	***	✓	✓	✓	✓	✓	✓		✓	125
	Moody's	0.027	0.076		0.220	0.071	***	✓	✓	✓	✓	✓	✓	✓		1874	0.220	0.071	***	✓	✓	✓	✓	✓	✓	✓		1874
GBP	ISS	-0.265	0.132	**	-0.218	0.101	**		✓	✓	✗	✓	✓	✓	✓	1160	-0.239	0.101	**		✓	✓	✓	✓	✓	✓	✓	997
	MSCI	-0.153	0.100		-0.318	0.136	**	✓		✓	✗	✓	✓	✓	✓	225	0.017	0.141		✓		✓	✓	✓	✓	✓	✓	239
	Refinitiv	-0.122	0.109		-0.309	0.132	**	✓	✓		✗	✓	✓	✓	✓	758	-0.328	0.132	**	✓	✓		✓	✓	✓	✓	✓	651
	RepRisk	0.394	0.090	***	0.194	0.138		✓	✓	✓		✓	✓	✓	✓	188	0.194	0.138		✓	✓	✓		✓	✓	✓	✓	188
	S&P Global	-0.035	0.115		-0.313	0.133	**	✓	✓	✓	✗		✓	✓	✓	863	-0.324	0.133	**	✓	✓	✓	✓		✓	✓	✓	740
	Sustainalytics	-0.025	0.128		-0.457	0.133	***	✓	✓	✓	✗	✓		✓	✓	311	-0.340	0.135	***	✓	✓	✓	✓	✓		✓	✓	273
	TVL	0.087	0.078		0.705	0.127	***	✓	✓	✓	✓	✓	✓		✓	142	0.705	0.127	***	✓	✓	✓	✓	✓	✓		✓	142
	Moody's	-0.142	0.122		-0.252	0.133	*	✓	✓	✓	✗	✓	✓	✓		883	-0.330	0.131	***	✓	✓	✓	✓	✓	✓	✓		786
JPY	ISS	0.134	0.071	*	0.143	0.056	***		✓	✗	✓	✓	✗	✓	✓	1680	0.187	0.055	***		✓	✓	✓	✓	✗	✗	✓	1737
	MSCI	0.092	0.057		0.477	0.068	***	✓		✓	✓	✓	✓	✓	✗	686	0.293	0.068	***	✓		✗	✗	✓	✓	✗	✓	1088
	Refinitiv	0.187	0.057	***	0.104	0.071		✓	✓		✓	✓	✗	✓	✓	2065	0.128	0.070	*	✓	✓		✓	✓	✓	✓	✓	1790
	RepRisk	0.096	0.063		0.063	0.074		✓	✓	✗		✓	✗	✓	✓	298	-0.327	0.058	***	✗	✗	✓		✓	✗	✓	✗	474
	S&P Global	0.093	0.074		0.330	0.068	***	✓	✓	✓	✓		✓	✓	✗	1769	0.173	0.070	**	✓	✗	✓	✗		✗	✗	✓	4156
	Sustainalytics	0.258	0.075	***	0.473	0.069	***	✓	✓	✓	✓	✓		✓	✗	591	0.333	0.070	***	✓	✓	✓	✓	✓		✓	✓	566
	TVL	0.033	0.054		0.707	0.072	***	✓	✓	✓	✓	✓	✗		✓	158	0.730	0.072	***	✓	✓	✓	✓	✓	✓		✓	136
	Moody's	0.047	0.060		0.265	0.068	***	✓	✓	✓	✓	✓	✓	✓		3230	0.265	0.068	***	✓	✓	✓	✓	✓	✓	✓		3230
USD	ISS	0.069	0.055		0.021	0.041			✓	✓	✓	✓	✓	✓	✓	3637	0.021	0.041			✓	✓	✓	✓	✓	✓	✓	3637
	MSCI	0.067	0.041		0.149	0.057	***	✓		✓	✓	✓	✓	✓	✓	1168	0.149	0.057	***	✓		✓	✓	✓	✓	✓	✓	1168
	Refinitiv	0.064	0.055		0.032	0.057		✓	✓		✓	✓	✓	✓	✓	3361	0.032	0.057		✓	✓		✓	✓	✓	✓	✓	3361
	RepRisk	0.022	0.056		0.036	0.057		✓	✓	✓		✓	✓	✓	✗	443	0.062	0.057		✓	✓	✓		✓	✓	✓	✓	381
	S&P Global	0.012	0.045		0.083	0.057		✓	✓	✓	✓		✓	✓	✓	3106	0.083	0.057		✓	✓	✓	✓		✓	✓	✓	3106
	Sustainalytics	0.140	0.069	**	0.168	0.057	***	✓	✓	✓	✓	✓		✓	✓	592	0.168	0.057	***	✓	✓	✓	✓	✓		✓	✓	592
	TVL	0.051	0.046		0.294	0.057	***	✓	✓	✓	✓	✓	✓		✓	349	0.294	0.057	***	✓	✓	✓	✓	✓	✓		✓	349
	Moody's	-0.027	0.056		0.098	0.057	*	✓	✓	✓	✓	✓	✓	✓		4300	0.098	0.057	*	✓	✓	✓	✓	✓	✓	✓		4300

Table 5. 2-month stock returns and ESG ratings. This table reports estimates of β from the OLS regression (27) and the 2SLS regression (29). The first set of columns shows OLS estimates, while the second and third set of columns show 2SLS coefficients from the Pruning and Lasso procedures, resp., described in Section 3.3. All returns are monthly and all reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. The checkmarks in the columns titled "Valid IV" indicate a selection of instruments that passes the Sargan-Hansen OIR test. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		OLS			Pruned IV												Lasso IV											
Region	Rater	Coeffs	StdErr		Coeffs	StdErr		Valid IV								Ftest	Coeffs	StdErr		Valid IV								Ftest
EUR	ISS	0.155	0.072	**	0.341	0.051	***		✓	✗	✗	✓	✓	✓	✗	1290	0.076	0.067			✗	✓	✗	✓	✗	✗	✓	2681
	MSCI	0.135	0.051	***	0.332	0.070	***	✓		✗	✓	✓	✓	✓	✗	1087	0.114	0.067	*	✗		✓	✗	✗	✓	✗	✓	2073
	Refinitiv	0.009	0.065		0.304	0.071	***	✓	✓		✗	✓	✓	✓	✗	1884	0.235	0.072	***	✓	✓		✗	✓	✓	✗	✓	2088
	RepRisk	0.145	0.068	**	0.446	0.052	***	✗	✓	✓		✗	✓	✓	✓	505	0.203	0.063	***	✗	✗	✓		✓	✓	✓	✗	636
	S&P Global	0.116	0.083		0.065	0.066		✗	✗	✓	✗		✓	✓	✓	1780	0.039	0.066		✗	✗	✓	✗		✗	✗	✓	3495
	Sustainalytics	0.125	0.077		0.422	0.071	***	✓	✓	✗	✓	✓		✓	✗	927	0.239	0.052	***	✗	✓	✓	✓	✓		✓	✓	1014
	TVL	0.036	0.066		0.753	0.071	***	✓	✓	✗	✓	✓	✓		✗	167	0.648	0.072	***	✓	✓	✓	✓	✗	✓		✓	143
	Moody's	0.028	0.078		0.269	0.072	***	✓	✓	✗	✓	✓	✓	✓		2001	0.241	0.072	***	✓	✓	✓	✗	✓	✓	✗		2533
GBP	ISS	-0.261	0.129	**	-0.231	0.099	**		✓	✓	✗	✓	✓	✓	✓	1142	-0.231	0.099	**		✓	✓	✗	✓	✓	✓	✓	1142
	MSCI	-0.170	0.098	*	-0.117	0.107		✗		✓	✗	✓	✓	✓	✓	221	-0.394	0.132	***	✓		✗	✗	✗	✓	✗	✗	517
	Refinitiv	-0.120	0.105		-0.329	0.129	***	✓	✓		✗	✓	✓	✓	✓	743	-0.325	0.129	***	✓	✗		✗	✓	✓	✓	✓	891
	RepRisk	0.398	0.090	***	0.163	0.103		✗	✓	✓		✓	✓	✓	✓	215	0.201	0.134		✓	✓	✓		✓	✓	✓	✓	186
	S&P Global	-0.058	0.115		-0.315	0.129	**	✓	✓	✓	✗		✓	✓	✓	846	-0.315	0.129	**	✓	✓	✓	✗		✓	✓	✓	846
	Sustainalytics	-0.003	0.127		-0.487	0.129	***	✓	✓	✓	✗	✓		✓	✓	306	-0.546	0.128	***	✓	✓	✗	✗	✓		✗	✓	426
	TVL	0.093	0.074		0.916	0.121	***	✓	✗	✓	✓	✓	✓		✓	151	-	-										
	Moody's	-0.157	0.122		-0.265	0.130	**	✓	✓	✓	✗	✓	✓	✓		863	-0.255	0.130	**	✓	✓	✓	✗	✓	✓	✗		1030
JPY	ISS	0.133	0.070	*	0.168	0.056	***		✓	✗	✓	✓	✗	✓	✓	1654	0.143	0.076	*		✗	✗	✗	✓	✗	✗	✓	3975
	MSCI	0.084	0.057		0.491	0.067	***	✓		✓	✓	✓	✓	✓	✗	675	-	-										
	Refinitiv	0.187	0.055	***	0.220	0.068	***	✓	✓		✓	✓	✗	✓	✗	1670	0.150	0.069	**	✓	✗		✗	✓	✗	✗	✓	3785
	RepRisk	0.089	0.060		-0.061	0.059		✗	✓	✗		✓	✗	✓	✓	364	-0.170	0.079	**	✗	✗	✗		✓	✗	✓	✗	594
	S&P Global	0.117	0.075		0.325	0.068	***	✓	✓	✓	✗		✗	✓	✗	2206	-	-										
	Sustainalytics	0.261	0.070	***	0.492	0.067	***	✓	✓	✓	✓	✓		✓	✗	591	0.335	0.055	***	✗	✓	✗	✓	✓		✗	✓	948
	TVL	0.038	0.053		0.681	0.071	***	✓	✓	✓	✓	✓	✗		✗	189	0.683	0.055	***	✗	✓	✓	✓	✓	✗		✗	236
	Moody's	0.060	0.060		0.250	0.067	***	✓	✓	✓	✓	✓	✗	✓		3609	0.243	0.067	***	✓	✓	✓	✗	✓	✗	✗		5390
USD	ISS	0.082	0.057		0.093	0.041	**		✓	✓	✓	✓	✗	✓	✗	3874	0.037	0.041			✓	✓	✓	✓	✗	✓	✓	4157
	MSCI	0.067	0.042		0.202	0.058	***	✓		✓	✓	✓	✓	✓	✗	1268	0.059	0.059		✓		✗	✗	✗	✗	✗	✓	2481
	Refinitiv	0.066	0.054		-0.004	0.041		✗	✓		✓	✓	✗	✓	✓	4293	0.009	0.059		✓	✗		✓	✓	✗	✗	✓	5553
	RepRisk	0.053	0.057		-0.049	0.059		✓	✓	✓		✓	✗	✓	✗	490	0.004	0.041		✗	✓	✓		✓	✓	✓	✗	503
	S&P Global	0.011	0.045		0.139	0.058	**	✓	✓	✓	✓		✓	✓	✗	3253	0.052	0.058		✓	✗	✗	✗		✗	✗	✓	8960
	Sustainalytics	0.140	0.072	*	0.256	0.059	***	✓	✓	✓	✓	✓		✓	✗	652	0.185	0.059	***	✓	✓	✓	✓	✓		✓	✓	581
	TVL	0.052	0.046		0.367	0.058	***	✓	✓	✓	✓	✓	✗		✓	397	0.352	0.058	***	✓	✓	✓	✓	✓	✓		✓	341
	Moody's	-0.018	0.056		0.097	0.058		✓	✓	✓	✓	✓	✗	✓		4880	0.105	0.058	*	✓	✓	✓	✓	✓	✓	✓		4223

Table 6. 3-month stock returns and ESG ratings. This table reports estimates of β from the OLS regression (27) and the 2SLS regression (29). The first set of columns shows OLS estimates, while the second and third set of columns show 2SLS coefficients from the Pruning and Lasso procedures, resp., described in Section 3.3. All returns are monthly and all reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. The checkmarks in the columns titled "Valid IV" indicate a selection of instruments that passes the Sargan-Hansen OIR test. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

Region	Rater	OLS Coeffs	OLS StdErr		Pruned IV Coeffs	Pruned IV StdErr		Pruned IV Valid IV								Pruned IV Ftest	Lasso IV Coeffs	Lasso IV StdErr		Lasso IV Valid IV								Lasso IV Ftest
EUR	ISS	0.159	0.073	**	0.079	0.067			✗	✓	✗	✓	✗	✓	✓	2000	0.071	0.066			✗	✓	✗	✓	✗	✗	✓	2618
	MSCI	0.145	0.052	***	0.339	0.071	***	✓		✗	✓	✓	✓	✓	✗	1073	0.117	0.067	*	✗		✓	✗	✗	✓	✗	✓	2043
	Refinitiv	0.002	0.065		0.310	0.072	***	✓	✓		✗	✓	✓	✓	✗	1851	-	-										
	RepRisk	0.142	0.068	**	0.398	0.062	***	✗	✗	✓		✗	✓	✓	✓	560	-0.061	0.065		✗	✗	✓		✓	✗	✓	✗	402
	S&P Global	0.111	0.084		0.034	0.066		✗	✗	✓	✗		✗	✓	✓	2301	0.035	0.066		✗	✗	✓	✗		✗	✗	✓	3451
	Sustainalytics	0.132	0.076	*	0.429	0.072	***	✓	✓	✗	✓	✓		✓	✗	916	-	-										
	TVL	0.041	0.065		0.759	0.072	***	✓	✓	✗	✓	✓	✓		✗	164	0.803	0.071	***	✓	✓	✗	✓	✗	✗		✓	195
	Moody's	0.031	0.079		0.311	0.072	***	✓	✓	✗	✗	✓	✓	✓		2275	-	-										
GBP	ISS	-0.271	0.127	**	-0.281	0.099	***		✓	✓	✗	✓	✓	✓	✓	1126	-0.281	0.099	***		✓	✓	✗	✓	✓	✓	✓	1126
	MSCI	-0.178	0.099	*	-0.182	0.110		✗		✓	✗	✓	✓	✓	✓	217	-0.432	0.129	***	✓		✗	✗	✗	✓	✗	✗	509
	Refinitiv	-0.151	0.108		-0.455	0.126	***	✓	✓		✗	✗	✓	✓	✓	740	-0.364	0.126	***	✓	✗		✗	✓	✓	✓	✓	875
	RepRisk	0.391	0.084	***	0.494	0.110	***	✗	✗	✓		✓	✓	✓	✓	199	0.193	0.104	*	✗	✓	✓		✓	✓	✓	✓	214
	S&P Global	-0.093	0.115		-0.385	0.126	***	✓	✗	✓	✗		✓	✓	✓	987	-0.364	0.126	***	✓	✓	✓	✗		✓	✓	✓	830
	Sustainalytics	-0.020	0.127		-0.525	0.126	***	✓	✓	✓	✗	✓		✓	✓	300	-0.596	0.126	***	✓	✓	✗	✗	✓		✗	✓	417
	TVL	0.093	0.076		0.973	0.120	***	✓	✗	✓	✓	✗	✓		✓	172	-	-										
	Moody's	-0.188	0.129		-0.396	0.128	***	✓	✓	✓	✗	✗	✓	✓		785	-0.297	0.127	**	✓	✓	✓	✗	✓	✓	✗		1008
JPY	ISS	0.122	0.069	*	0.170	0.054	***		✓	✗	✓	✓	✗	✓	✓	1629	0.149	0.076	*		✗	✗	✗	✓	✗	✗	✓	3917
	MSCI	0.073	0.056		0.409	0.067	***	✓		✓	✓	✓	✗	✓	✗	707	-	-										
	Refinitiv	0.181	0.054	***	0.160	0.067	**	✓	✓		✗	✓	✗	✓	✓	2294	0.153	0.068	**	✓	✗		✗	✓	✗	✗	✓	3762
	RepRisk	0.085	0.061		0.175	0.057	***	✗	✓	✗		✗	✗	✓	✓	303	-	-										
	S&P Global	0.114	0.075		0.304	0.067	***	✓	✓	✓	✗		✗	✓	✗	2185	-	-										
	Sustainalytics	0.261	0.068	***	0.468	0.066	***	✓	✓	✓	✓	✓		✓	✗	590	0.326	0.054	***	✗	✓	✗	✓	✓		✗	✓	946
	TVL	0.040	0.053		0.646	0.070	***	✓	✓	✓	✓	✓	✗		✗	188	0.647	0.054	***	✗	✓	✓	✓	✓	✗		✗	235
	Moody's	0.067	0.059		0.237	0.066	***	✓	✓	✓	✓	✓	✗	✓		3581	0.231	0.066	***	✓	✓	✓	✗	✓	✗	✗		5348
USD	ISS	0.082	0.057		-0.001	0.056			✗	✗	✓	✓	✗	✓	✓	5636	0.014	0.053			✗	✓	✗	✓	✗	✗	✓	7896
	MSCI	0.069	0.041		0.214	0.059	***	✓		✓	✓	✗	✓	✓	✗	1455	-	-										
	Refinitiv	0.060	0.054		-0.018	0.056		✗	✗		✓	✓	✗	✓	✓	5185	-0.013	0.046		✗	✗		✗	✓	✗	✗	✓	9622
	RepRisk	0.066	0.056		0.201	0.046	***	✗	✗	✗		✓	✗	✓	✓	509	-0.048	0.053		✗	✗	✓		✓	✗	✓	✗	761
	S&P Global	0.011	0.046		0.117	0.058	**	✓	✓	✓	✓		✗	✓	✗	3705	-	-										
	Sustainalytics	0.135	0.073	*	0.353	0.060	***	✓	✓	✓	✓	✗		✓	✗	676	0.193	0.041	***	✗	✓	✓	✗	✓		✗	✓	947
	TVL	0.051	0.046		0.367	0.059	***	✓	✓	✓	✓	✓	✗		✗	463	0.351	0.058	***	✓	✓	✓	✓	✓	✓		✗	387
	Moody's	-0.019	0.055		0.113	0.058	*	✓	✓	✓	✓	✗	✗	✓		5281	0.097	0.058		✓	✓	✓	✗	✓	✗	✗		7171

Notice further that the number of rejected instruments grows as we increase the horizon over which we measure returns. (The F-statistics from the first stage are still very high and in fact go up, so this drop in the number of valid instruments does not hinder our estimation.) The most likely reason for this is that the ESG scores of some raters are updated infrequently, e.g., once a year and not at the same time, and other raters (e.g., Truvalue Labs) update their scores daily. If scores of different raters contain correlated measurement errors — e.g., because they use similar models or data imputation methods, or because they incorporate information from the same unjustified media attention episode — we are more likely to observe this over a longer horizon. That is a likely reason for having more OIR rejections in Tables 5 and 6 relative to Table 4.

One may wonder whether the effect and significance of our results come from a specific month. For that reason, we rerun our estimations for month-by-month returns for each of the months $t+1$ to $t+8$ on the left-hand side. The results can be found in Tables A4–A6 in the Appendix. It is evident from the tables that our results for each separate month closely resemble each other in terms of the magnitudes of the coefficients and their standard errors. Overall, the coefficients increase on average by a factor of 2.6 for all regions, raters, and each of the months $t+1$ to $t+8$.

We further explore the stability of the attenuation bias by varying the returns' horizon from 1 to 8 months. The results are summarized in Figure 1, while the details are provided in Appendix A.3. The black line indicates the average ratio of the coefficient for the h-month returns to the coefficient for the 1-month returns (for the same rater and region). It shows that the OLS coefficients increases almost monotonically for longer horizons. The average normalized OLS estimate with (monthly) 8-month returns is 1.39, implying that the coefficients increase by 40%. The blue line denotes the average ratio of the Pruned IV coefficients for h-month returns to the OLS coefficient for 1-month returns. This statistic indicates that the increase in the instrumented coefficients is higher relative to that for the OLS coefficients. The blue envelope represents the 5 to 95 percent confidence interval. The dashed yellow line and yellow envelope show the equivalent statistics for the Lasso IV procedure. The plot illustrates that both IV procedures produce the increase in coefficients for longer time horizons, which indicates that the attenuation bias is stable even when changing the left-hand-side variable. The average normalized Pruned estimates start at 2.68 when the horizon is only one month. The ratio between the instrumented coefficients and the OLS coefficients fluctuates between 2.68 and 3.26. Given the confidence bands, any test of equality is likely to not be rejected. For the Lasso estimates, the patterns are very similar. At one month, the same ratio is 2.69. The relative coefficients fluctuate between 2.21 (for 8-month returns) and 2.92 (2-month returns). Finally, the confidence intervals also track the upwards movement observed for OLS, and they overlap between the Lasso and the Pruned

IV estimates at all horizons.[16] One possible reason for the positive trend in the coefficients is that the true β indeed changes with the horizon over which stock returns are measured. Another reason is that omitted variable bias (the term with the covariance $cov(Y_{k,t}, M_{h,k,t})$ in (11)–(12)) increases with horizon h.

Thus far, we have explored how the OLS and 2SLS coefficients change as we vary the horizon over which we measure returns. Although the coefficients vary, their ratio should not (Equation (13)). Our next step is to explore this ratio and to relate it to the implied noise in ESG ratings.

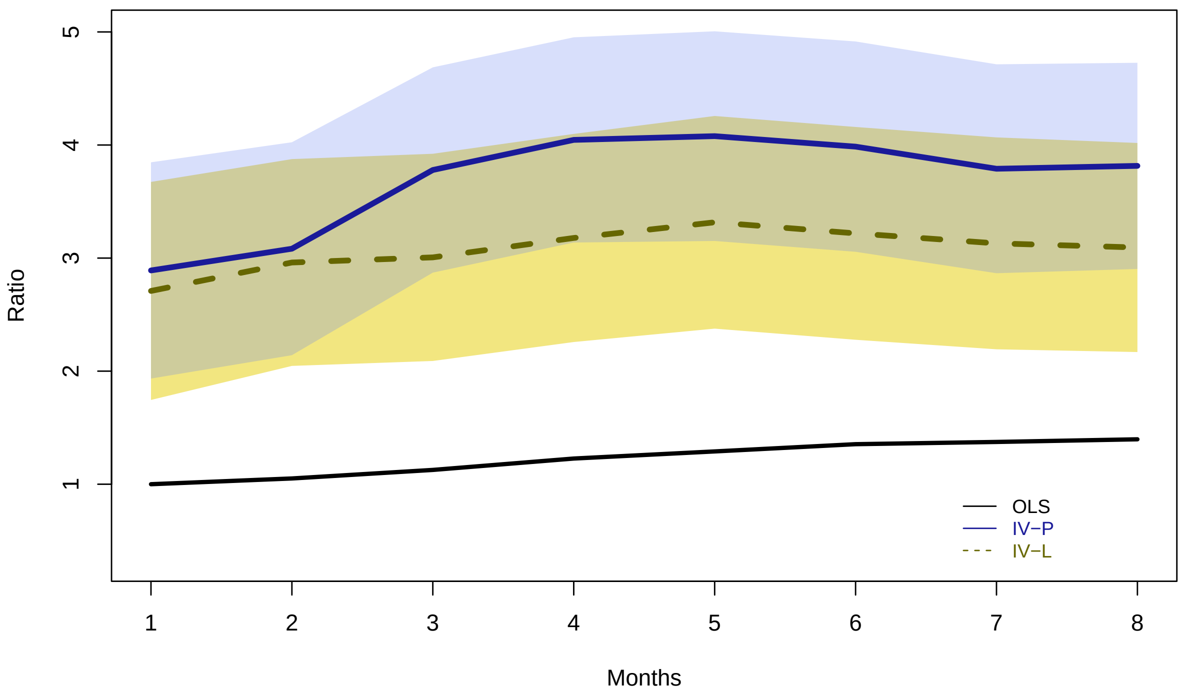


Figure 1. OLS and 2SLS coefficients for different return horizons. This figure shows the average ratio of the OLS coefficient for the h-month return specification to the OLS coefficient for the 1-month return one (black curve), as well as the average ratio of the 2SLS coefficient for the h-month return specification to the OLS coefficient for the 1-month return one for both Pruned IV (dark blue curve) and Lasso IV (dashed green curve). On the horizontal axis, we vary the horizon over which we measure returns from 1 to 8 months. The averages are computed over raters and regions. The black line is for the OLS coefficients, the blue line is for the Pruned IV coefficients, and the dashed green for the Lasso IV. The blue envelope represents the 5% to 95% confidence interval for the Pruned IV estimate and the yellow envelope shows the equivalent for the Lasso IV procedure.

[16]This should not be interpreted as a test because these are different coefficients and the testing should occur at the individual horizon, rater, and region. See Appendix A.3 for the individual estimates for each region, rater, and horizon and their standard deviations.

5.1 Implied Noise in ESG Ratings

Equation (6) derives κ_i, the noise-to-signal ratio, which allows us to calculate the noise that is implied by the difference between the OLS and the 2SLS coefficients for each rater i. The overall average of the noise-to-signal ratio is 61.7% across all 8 raters, 4 regions, 8 return horizons, and the two 2SLS estimation procedures, Pruned IV and Lasso IV. Figure 2 reports how the implied noise-to-signal ratios vary by geographical region, rater, and horizon.

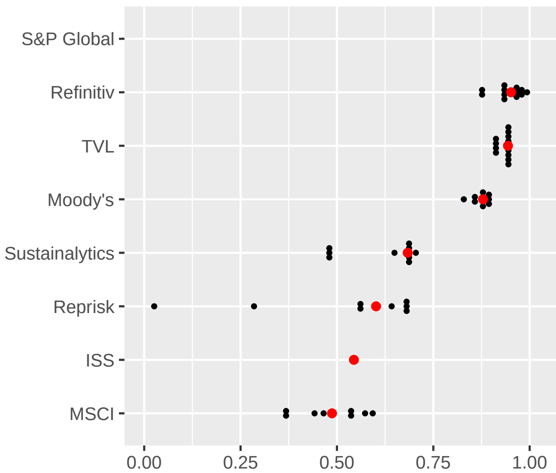
Figure 2 shows four panels, one for each region. Within each panel, raters are ranked by their median κ_i, shown as a red dot. Around the median, a histogram of the remaining estimates is shown with black dots. Note that Equation (6) can only be used only when there is attenuation bias in the OLS coefficient as it implies negative variances otherwise. As a result, we do not have the same number of values for each rater.

Figure 2 offers three insights. First, in most cases, different estimates of κ_i for the same rater i are close to each other, visible as distinct groups. This is another way of illustrating that the attenuation bias remains relatively constant for different time horizons and estimation techniques. Second, there are marked differences between raters. ISS has the lowest noise-to-signal ratio in Japan, U.K., and the U.S., although we should acknowledge that in the U.S. this rank relies on just one single estimate. MSCI has the lowest median κ in the eurozone. Reprisk also tends to have low κ values across regions. Truvalue Labs and S&P Global tend to have high noise-to-signal ratios. As discussed earlier, these two ratings have some unique methodological features and are most frequently accepted as instruments. The reason is, most likely, that the noise in these ratings is truly orthogonal to that in other ratings, satisfying our identifying assumptions (7), (8), and (9).

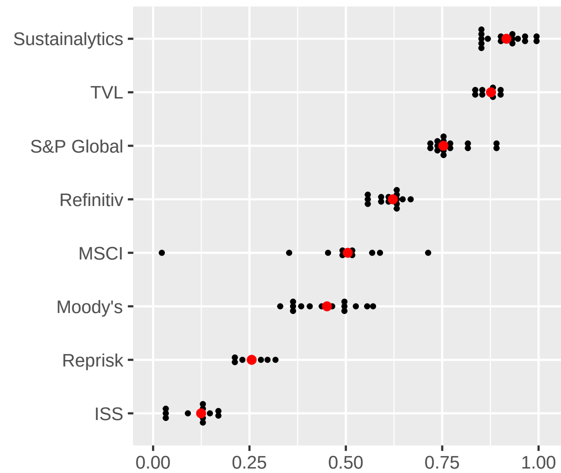
Third, the ranking changes across regions. This indicates that the question of which rater has the lowest k also depends on the sample, which in turn implies that users of ESG ratings who wish to use our methodology should run our procedure on their specific sample and regression specification to ascertain the level of noise they are dealing with.

By examining Figure 2, one might hastily conclude that one should source scores only from ESG raters with the most precise measurement, located at the bottom of the figure. We caution against such conclusion. Disregarding scores of other raters amounts to discarding valuable information about the unobservable ESG performance the ratings are trying to measure. Intuitively, by combining different ratings, and in particular ratings that rely on different information sources and which contain different sorts of noise, one can get the most precise signal about the unobservable ESG performance.

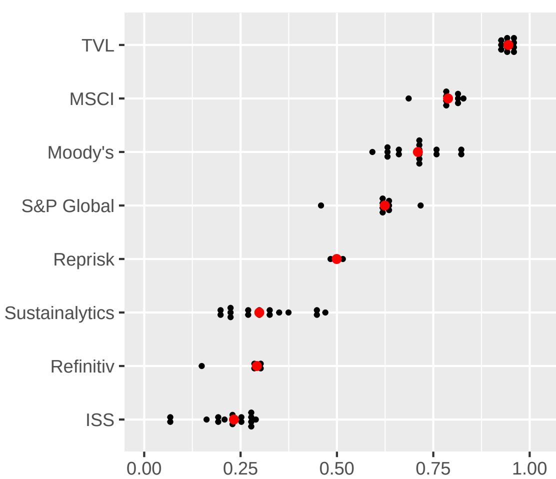
How often do our estimations fail to produce an estimate or to detect an attenuation bias? Table A7 in the Appendix presents the big-picture outcome from 512 specifications we



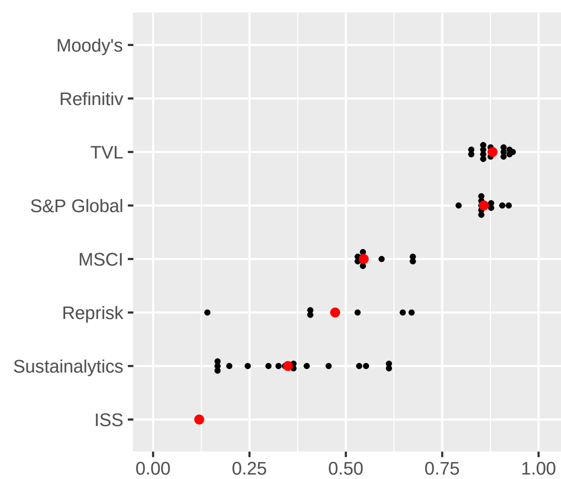
(a) Eurozone



(b) U.K.



(c) Japan



(d) U.S.

Figure 2. Implied noise in ESG ratings. This figure shows the implied noise in ESG scores κ_i, as defined in Equation (6), for each rater i. Panel (a) shows the data for the eurozone, Panel (b) for the U.K., Panel (c) for Japan, and Panel (d) for the U.S.. The black dots show a histogram on the horizontal axis with a bin width of 0.015 for the values estimated for different return horizons on the left-hand side (1 to 8 months) for both the Pruned IV and Lasso IV procedures. There is a maximum of 16 values per rater and region, but the implied noise cannot be computed when the 2SLS coefficients are smaller than the OLS coefficients or when the set of valid instruments is empty. The red dot indicates the median of the available values. Raters are sorted by this median value within region.

consider (by varying region, return horizon, rater, and the 2SLS estimation procedure). Out of a total of 512 possible estimates, we are able to estimate 427. This means that we found 427 possible estimates in which the overidentifying restriction was *not* rejected (83.7%).

Most of the coefficients that we could not estimate use the Lasso IV procedure. The Pruned IV procedure produces an estimate in 98% of the cases, compared to 68.4% for the Lasso IV procedure. Of these 427 coefficients, 302 exhibit attenuation bias (70.7%). Of those that do not, 33 (7.7%) of the coefficients switch sign (OLS relative to the corresponding 2SLS) and for 92 (21.5%) of the estimates the 2SLS coefficient is smaller in absolute value than the corresponding OLS one. Thus, in more than two-thirds of the cases where estimation is feasible, we observe attenuation. Table A7 in the Appendix contains further details regarding the quality of our estimation.

6 Simulations

In this section, we present a series of simulations to test the robustness of our IV estimation strategy. First, we compare the IV-based procedure to alternative noise-reduction procedures that are frequently used by practitioners: averaging the rating scores and using the principal component analysis. We demonstrate that, in our application, the IV procedure is superior to the other two. Second, we show that the OIR test is highly sensitive to simulated violations of our identifying assumptions for 2SLS estimation, providing reassurance that the selection for valid instruments is robust. Third, we explore the case of multiple noisy indicators to the point where there are more sources of noise than instruments, as well as noise coming from measurement in indicators versus noise coming from the aggregation of indicators. We can show that an IV estimation based on multiple instruments tends to come close to the true coefficients in all those settings.

6.1 Our Estimation Procedure versus Alternative Noise-Reduction Techniques

For the simulations, we generate 15,000[17] observations of stock returns and ESG performance from the following structural model:

$$r_k = \beta \cdot Y_k + \epsilon_k, \tag{30}$$

where k indexes observations. Y_k is a normal random variable with a mean of 0 and a standard deviation of 1. The coefficient of interest is β, which we set equal to 0.5. We assume that

[17]This is in line with the average number of observations in our empirical setup, where we have 65,307 firm-month observations across the four regions.

there are N rating agencies indexed by i, with scores modeled as in our preceding analysis:

$$s_{k,i} = Y_k + \eta_{k,i}. \tag{31}$$

The focus is on rating $s_{k,1}$, which is our "problematic" regressor, i.e., it is measured with noise. For the remainder of this section, we suppress the subscript k for expositional clarity.

We perform two simulations. In the first, we generate $N = 3$ ratings and benchmark the 2SLS procedure to two alternative noise-reduction approaches, simple averaging and principal component analysis. Here, all instruments are valid, i.e., satisfy assumptions (7), (8), and (9). In the second simulation, we add an invalid instrument s_4, whose errors η_4 correlate with η_1, thereby violating assumption (9).

In each simulation, we compare several procedures. First, we run a simple OLS:

$$r = \alpha + \beta \cdot s_1 + \nu. \tag{32}$$

Second, one may conjecture that an index constructed as an average of the raters' scores would be less noisy than each rating individually. We therefore construct a simple average of the rating scores, $s^{avg} = 1/N \sum_{i=1}^{N} s_i$, and estimate the following regression:

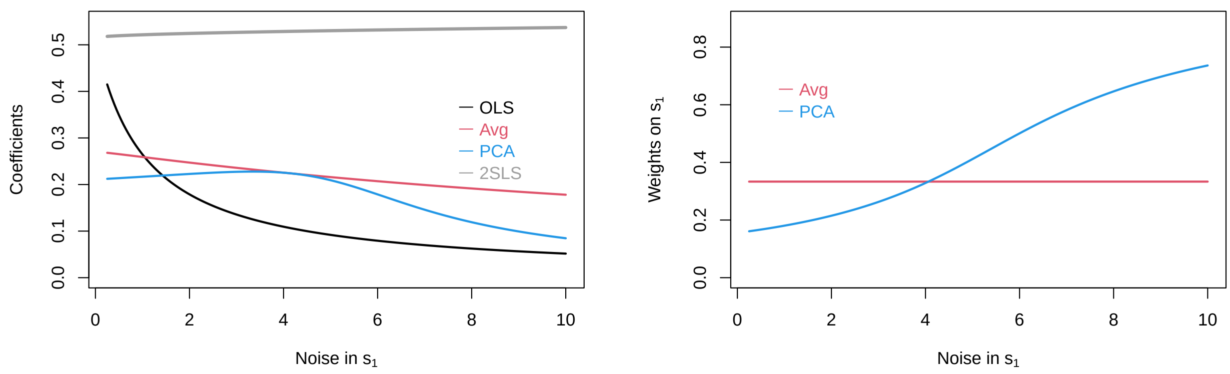
$$r = \alpha_{avg} + \beta_{avg} \cdot s^{avg} + \varepsilon. \tag{33}$$

Third, one may suggest using the principal components analysis as a noise reduction procedure. We therefore estimate the principal component of the N rating agencies' scores, denoted as s^{pc}, and estimate following the regression:

$$r = \alpha_{pc} + \beta_{pc} \cdot s^{pc} + \varepsilon. \tag{34}$$

Finally, we perform our instrumental variable estimation and the corresponding OIR test, using the ratings $i = \{2, \ldots, N\}$ as instruments for s_1 in the first stage.

The first simulation varies σ_{η_1}, the variance of the noise in s_1, from 0.25 to 10. The variance of the noise for s_2 and s_3 is kept constant at a value of 4. For each value of σ_{η_1}, we perform the above noise-reduction procedures and summarize the estimation results in Figure 3. Panel (a) plots the estimate of β for the OLS, simple average, PCA, and 2SLS estimation methods. The 2SLS estimate, shown in gray, is very close to the true β of 0.5 for any level of noise in s_1. It varies from 0.518 to 0.537 over the range of simulated noise in s_1. The OLS estimate displays attenuation bias, which, as expected, increases with the level



(a) Attenuation bias

(b) Weight on s_1

Figure 3. Simulation 1: Comparison of OLS, 2SLS, simple average, and PCA. This figure plots the estimates of β from (32), (33), (34), and our 2SLS procedure (Panel (a)) and the weights a given estimation procedure puts on the rating s_1 (Panel (b)), for varying levels of noise in the rating s_1. All instruments are valid.

of noise in s_1. The estimate based on the simple average is inferior to OLS for low levels of noise. As the noise in s_1 increases, it becomes advantageous to include scores of other raters. This is true even when the other scores are noisier than the regressor, because the error terms are independent. The estimate based on the first principle component is never better than the simple average which is because the PCA finds the linear combination of raters' scores that maximizes the observed variance. This approach is quite useful when the variables are measured correctly; however, if the variables are measured with noise, the noise becomes part of the observed variance and a PCA puts greater weight on noisier ratings. This pattern is apparent from Panel (b) of Figure 3. Simple average puts the same weight on all ratings, regardless of their noisiness. Intuitively, one should put a lower weight on a noisier rating. This is, in fact, what our 2SLS procedure does, as noisier indicators have smaller coefficient estimates in the first stage.

Our second simulation adds one instrument, s_4, which is invalid by construction. Specifically, we vary the correlation between η_1 and η_4 from -0.5 to 0.5. As a result, s_1 and s_4 are related not only through Y but also through their errors. The standard deviations of the noise, σ_{η_i}, $i = 1, 2, 3, 4$, are set to be $\{3, 1, 1, 1\}$, respectively. This reflects the case where the potential instruments are less noisy than the regressor.[18] Table 7 presents the results of this simulation.

Let us first concentrate on the row corresponding to the correlation of zero. For this row, the assumptions of the classical errors-in-variables problem are satisfied. As before, we clearly see the attenuation bias in the OLS estimation (32): an estimate of 0.060 instead of 0.5. The bias becomes smaller if we use the simple average s^{avg} (33) as a regressor: the coefficient increases to 0.307. Despite the fact that the other three scores, s_2, s_3, and s_4, contain less noise than s_1, the resulting estimate is still far from 0.5. The third column shows the estimates when we use the first principal component s^{pc} (34) as the regressor. With 0.072, the estimate is nearly as biased as the OLS estimate. The next two columns show the estimates from the 2SLS procedures, the first one including all available instruments, $Z = \{s_2, s_3, s_4\}$, the second one just two instruments, $Z = \{s_2, s_3\}$. Notice that both estimates are very close to 0.5, as the correlation between errors is zero and thus all instruments are valid. The last two columns show the p-values of the OIR tests for the 2SLS estimations and both models pass the test.

Let us now turn to the rows in Table 7 where the correlations between the errors η_1 and η_4 are different from zero. The OLS estimate does not change, since s_4 is not used in this regression. The estimate of the coefficient on the average s^{avg} in column (2), however, increases for negative correlations and decreases for positive correlations. This is because negatively correlated errors cancel each other out. As a result, the average is becoming less noisy, and the coefficient is moving slightly towards the true value of 0.5. The coefficient on the first principal component in column (3) varies because the change in the correlation of the scores implies small changes in the eigenvector. However, all estimated coefficients remain far from the true value of 0.5.

The estimates of the 2SLS regressions in column (4) highlight the outcome of our estimation procedure for the case in which an invalid instrument is present. The estimated coefficient moves away quickly from the true value of 0.5. The coefficients in column (4) are close to 0.5 only when the correlation is between -0.1 and 0.1. The 2SLS estimates that exclude s_4 in column (5), however, recover the true coefficient for any level of correlation. Most importantly, the OIR test in column (6) shows that models with invalid instruments are rejected with high reliability. The OIR tests in column (7) for the set of valid instruments are never rejected. Furthermore, cases where the OIR test is not rejected (even though the

[18] The qualitative results do not depend on the choice of the standard deviations of the noise, except in the case where the noise on the regressor is exactly zero. The level of variance in the instruments changes the region where the OIR is rejected. The general message, though, remains the same. First, when the noise in instruments is sufficiently correlated, the OIR test is rejected; and second, when there is no rejection, even in the cases where the instruments are invalid, the point estimates of the 2SLS are very close to the true coefficient.

instruments are invalid) are instances in which the point estimates of the 2SLS procedure are very close to the true coefficient.

Table 7. Simulation 2: Estimates for the case where the measurement errors η_1 and η_4 are correlated. The true value of the coefficient is 0.5. Columns (1)-(3) report the estimates from regressions (32)–(34), respectively. 2SLS All reports the estimates with $Z = \{s_2, s_3, s_4\}$ and 2SLS Z reports the estimates with only $Z = \{s_2, s_3\}$, i.e., the subset of valid instruments. OIR stands for the Sargan-Hansen OIR test, and the corresponding column reports the p-values for the test.

Correlation of η_1 and η_3	(1) OLS	(2) Average	(3) PCA	(4) 2SLS All	(5) 2SLS Z	OIR All (p-value)	OIR Z (p-value)
-0.5	0.060	0.341	0.069	0.107	0.524	0.000	1.000
-0.4	0.060	0.334	0.071	0.177	0.524	0.000	1.000
-0.3	0.060	0.327	0.073	0.285	0.524	0.000	1.000
-0.2	0.060	0.320	0.075	0.422	0.524	0.000	1.000
-0.1	0.060	0.313	0.077	0.528	0.524	0.000	1.000
0	0.060	0.307	0.078	0.530	0.524	0.951	1.000
0.1	0.060	0.301	0.078	0.455	0.524	0.003	1.000
0.2	0.060	0.296	0.079	0.365	0.524	0.000	1.000
0.3	0.060	0.290	0.079	0.290	0.524	0.000	1.000
0.4	0.060	0.285	0.079	0.234	0.524	0.000	1.000
0.5	0.060	0.280	0.079	0.193	0.524	0.000	1.000

Figure 4 shows how the p-value from the OIR test evolves with the correlation between η_1 and η_4. Notice that any correlation larger than 0.1, in absolute value, produces p-values below the threshold. This simulation suggests that the OIR test has sufficient power to reject when an invalid instrument is present.

This result gives us confidence in the ability of our two procedures to select instruments. The Pruning procedure tends to accept a feasible set too soon. However, given how strong the rejections are in the simulation, even if the size of the test changes (e.g., from 0.05 to 0.005), we are likely to find a set of instruments that are valid. Of course, this is a conjecture, but it is also supported by our second procedure for selecting instruments, based on our Lasso procedure and described in Section 3.3, in which we tend to reject too early.

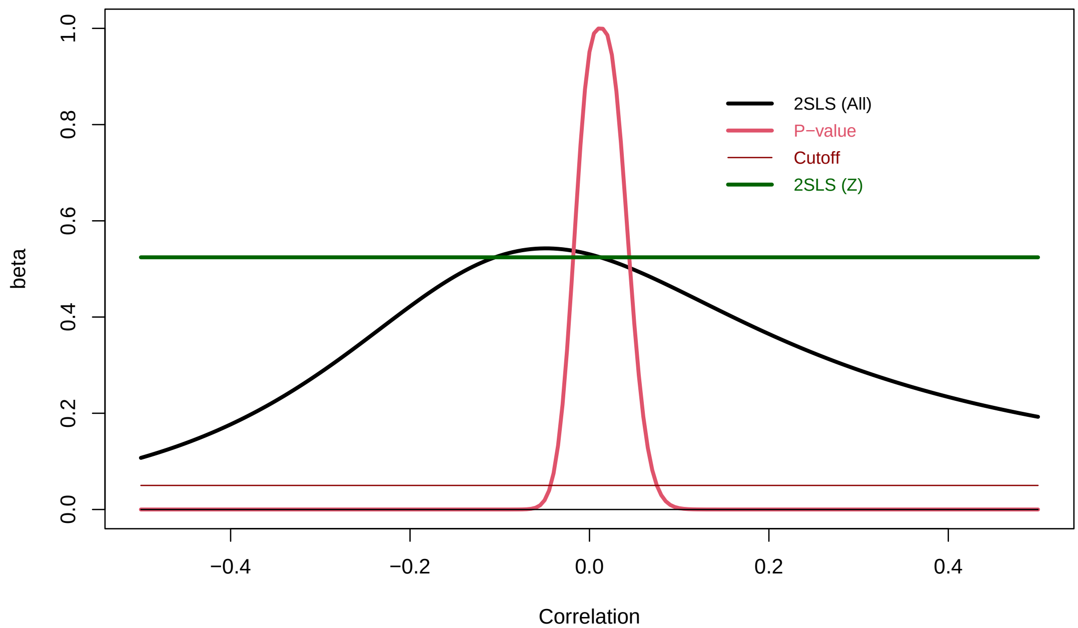


Figure 4. The power of the Sargan-Hansen OIR test. This figure plots the 2SLS coefficients from columns (4) and (5) of Table 7 as functions of the correlation between measurement errors in the simulated ESG ratings s_1 and s_4. The green line is the 2SLS estimate with valid instruments. The thick black line is the 2SLS estimate when including invalid instruments. The thick red line is the p-value of the Sargan-Hansen OIR test when invalid instruments are present, analogous to column (6) in Table 7. The thin red line represents the 0.05 p-value.

6.2 Aggregation of Many ESG Indicators

One potential criticism of our procedure is that ESG performance is a complicated aggregate, which different agencies define in different ways. For example, one agency may include, say, water pollution among the attributes it measures and another agency may not. Would our procedure still recover the true effect of ESG performance? This section presents a simulation that addresses this question.

As explained in detail in Appendix A.4, the rating agencies' scores are computed as a weighted average of many indicators, corresponding to disaggregated ESG attributes (e.g., carbon emissions, labor practices):

$$s_i = \sum_{a \in \{1,n\}} w_{a,i} \cdot I_{a,i}, \tag{35}$$

where i indexes ESG rating agencies, a indexes attributes that the agency considers, $I_{a,i}$ is rater i's measure of attribute a, and $w_{a,i}$ are the weights.

The true value of Y is given by a similar construct,

$$Y = \sum_{a \in \{1,n\}} w_a^\star \cdot I_a^\star,$$

where $I_a^\star$ are the true values of the indicators and $w_a^\star$ are the true weights—i.e., the weights that the representative ESG investor assigns to individual indicators, which reflect her preferences or social preferences.

The measurement error of each rating agency can be decomposed as follows:

$$s_i = Y + \underbrace{\sum_{a \in \{1,n\}} w_{a,i} \cdot \underbrace{(I_{a,i} - I_a^\star)}_{\eta_{I_{a,i}}} + \sum_{a \in \{1,n\}} \underbrace{(w_{a,i} - w_a^\star)}_{\eta_{w_{a,i}}} \cdot I_a^\star}_{\eta_{Y_i}}.$$

There are two sources of noise in this decomposition: the measurement error at the level of the indicator,

$$I_{a,i} = I_a^\star + \eta_{I_{a,i}}, \tag{36}$$

and the discrepancy in the weights,

$$w_{a,i} = w_a^\star + \eta_{w_{a,i}}. \tag{37}$$

The validity of instruments requires orthogonality of $\eta_{I_{a,i}}$ and $\eta_{w_{a,i}}$ across rating agencies. The measurement error in the aggregated rating, η_{Y_i}, parallels our η_i in Section 3 (see Equation (2)). Appendix A.4 provides more detail.

Even under the above assumptions, it is unclear how our 2SLS procedure performs when there are many sources of noise and a limited number of instruments. If there are 24 possible attributes, then there are 24 sources of measurement error and 23 different weights.[19] Would our procedure that relies on only 7 other raters' scores to use as instruments recover the true coefficient?

There are two sources of noise in rating agencies' scores that serve as our instruments. First, every individual attribute in (35) is measured with noise. For simplicity, in this simulation, we assume that measurement errors in each indicator have the same variance equal to 1. Second, each rating agency measures only a subset of the attributes. Therefore, the scores from the rating agencies that will be used as instruments present an incomplete

[19]The number 24 is not arbitrary. For example, the total number of the Sustainability Accounting Standards Board's (SASB) attributes is 26, 24 of which are in our data set.

picture of the true ESG performance. Such incomplete coverage, which Berg, Kölbel, and Rigobon (2020) term differences in scope, is a source of potential bias. Third, we assume that all agencies use equal weighting in their ESG scores. Hence, our simulation explores the effects of two sources of noise present in the ratings: (i) the noise with which individual indicators are measured and (ii) the noise resulting from the coverage problem.

We continue using the following structural model to simulate observations of stock returns:

$$r_k = \beta \cdot Y_k + \epsilon_k,$$

where k indexes observations, $Y_k = \frac{1}{n}\sum_{a\in\{1,n\}} I^{\star}_{a,k}$, and $\beta = 0.5$. The corresponding reduced-form model is

$$r_k = \alpha + \beta \cdot s_{1,k} + \nu_k. \tag{38}$$

We again estimate Equation (38), first by OLS and then by 2SLS, using other rating agencies' scores as instruments. We assume that there are 24 attributes and that these attributes are orthogonal to each other. We further assume that the rating of interest, s_1, includes all 24 attributes in its ESG rating and that it measures each one with noise. We assume that only 5 other rating agencies' ESG scores are available for use as instruments. As in our previous simulation, we generate 15,000 draws to work with. Errors with which each attribute is measured are classical and satisfy the identifying assumptions spelled out in Appendix A.4. These assumptions imply that all instruments are valid.

The results of the simulation are presented in Table 8, which has the following structure. In the first column, we report the coefficient from our baseline OLS regression (32). In the remaining columns, we present 2SLS estimates, in which we vary the number of instruments used in the first stage from 1 to 5, labeled as IV1 to IV5. In rows, we vary the number of attributes covered by each rating over the set of $\{1, 2, 4, 6, 12, 24\}$. In the first set of rows, the instruments cover only 1 of the possible 24 attributes, in the second set 2 of the possible 24, and so on, until the last set of rows in which each instrument covers all the attributes. We randomly choose the subset of attributes that is covered by each rater, with replacement. The rating is then computed as a simple average of the simulated indicators.

Let us now discuss the first set of (three) rows presented in Table 8, in which each instrument measures only one attribute. Notice that the OLS coefficient is 0.076. The 2SLS coefficient with one instrument is 2.096, and it falls to 1.448 when 5 instruments are used. The second row in the set presents the standard errors of the estimates. Notice that the precision of the estimates does not necessarily improve with the number of instruments. The OLS coefficient is not statistically different from zero and is statistically smaller than 0.5 at the 1 percent significance level. The coefficients in IV1 to IV5 estimations are consistently

Table 8. Simulation 3: Fewer instruments than sources of noise. The true value of the coefficient is 0.5. In the first column, we report the coefficient from our baseline OLS regression (32). In the remaining columns, we present 2SLS estimates, in which we vary the number of instruments used in the first stage from 1 to 5, labeled as IV1 to IV5. In rows, we vary the number of attributes covered by each rating. The rating s_1 that we are instrumenting covers all 24 attributes.

Number of Attributes per Rating		OLS	IV1	IV2	IV3	IV4	IV5
1	Coefficient	0.076	2.096	1.6	1.716	1.367	1.448
	Std Error	0.034	0.977	0.802	0.739	0.682	0.633
	1st stage F-stat		18	14	11	9	9
2	Coefficient	0.076	0.874	0.921	1.001	0.958	1.072
	Std Error	0.034	0.549	0.456	0.444	0.397	0.378
	1st stage F-stat		59	42	30	28	25
3	Coefficient	0.076	0.826	0.342	0.315	0.161	0.122
	Std Error	0.034	0.617	0.503	0.46	0.416	0.403
	1st stage F-stat		46	35	28	26	22
4	Coefficient	0.076	0.421	0.427	0.514	0.635	0.624
	Std Error	0.034	0.548	0.445	0.435	0.424	0.41
	1st stage F-stat		59	45	31	25	21
6	Coefficient	0.076	0.522	0.187	0.614	0.61	0.122
	Std Error	0.034	0.537	0.479	0.422	0.419	0.39
	1st stage F-stat		61	39	33	25	23
12	Coefficient	0.076	0.876	1.023	0.625	0.642	0.644
	Std Error	0.034	0.539	0.463	0.389	0.37	0.354
	1st stage F-stat		61	41	39	32	28
24	Coefficient	0.076	0.116	0.289	0.436	0.426	0.436
	Std Error	0.034	0.395	0.352	0.327	0.317	0.312
	1st stage F-stat		114	72	55	44	37

higher than their OLS counterpart, but, in this setting, they are also statistically distinct from the true parameter $\beta = 0.5$. The third row reports the F-statistic of the first stage, which indicates that the instruments still have relevance, which is expected given that all the scores are correlated through the fundamental ESG attributes. This indicates that in the extreme case where all instruments combined observe merely 5 of the 24 indicators that make up Y, a 2SLS approach still alleviates attenuation bias, but fails to come close to the true parameter.

However, this quickly improves when the ratings contain slightly more information about Y. When each rating covers 4 or more attributes, the IV estimates are not statistically different from the true parameter $\beta = 0.5$. We have also run OIR tests for all IV specifications that use two or more instruments, and observed no rejections. This is not surprising. While our instruments have incomplete coverage, they are still valid instruments, irrespective of whether the noise comes from the measurement of individual indicators or from an omitted fundamental.

In this section, we have studied a simulation in which ESG ratings include a large number ($n = 24$) of possible attributes. Estimation performed on more granular ratings, E, S, or

G, or, better yet, at an individual indicator level (measuring just one attribute) would reduce estimation error introduced by data aggregation. Ideally, we would like to have a separate instrument for each individual indicator included in a rating. Instead, we are using a weighted sum of indicators (e.g., an ESG score) as an instrument for another weighted sum (e.g., an ESG score of another rater). Appendix A.6 demonstrates how the simulation in this section changes if we perform our procedure at a more granular level, in which the scores are weighted averages of only $n = 8$ indicators, as opposed to $n = 24$. Table A8 in the Appendix reveals that our estimated coefficients are much closer to the true parameter $\beta = 0.5$.

7 Limitations

Our empirical analysis is not free of limitations. First, our time series are short, and especially so for stock return regressions. We can do very little about this problem because the rating agencies to date have produced data for a relatively short time frame. Our estimation relies primarily on cross-sectional variation because many rating agencies change their scores once a year at most, which implies that in practice the time series is even shorter than the time span of our sample.

Second, many rating agencies are in the process of consolidation, which often involves revising their procedures. Thus, some rating agencies back-fill their past scores based on a revised procedure. This is particularly problematic if the back-fill is based on stock-relevant information. If point-in-time scores are not available, practitioners and applied work can use our procedure to diagnose this problem, which will show up as a rejection of the OIR tests.

Third, the ESG score could be capturing unobservable firm characteristics. In our specification this corresponds to the omitted variable. As we have discussed before, the omitted variable has no impact on our results regarding the attenuation bias. However, it does change the structural interpretation of the coefficient. A potential omitted variable could be unexpected capital flows towards ESG. In the language of our model, we do not observe unexpected innovations to λ. According to our model, if the share of ESG investors is constant, the coefficients should be negative but when capital flows increase unexpectedly, the coefficient becomes positive. Another potential omitted variable could be management quality. Better managers might be able to foster more collaborative work environments that result in less labor mistreatment at the same time. This could impact both stock returns and ESG performance. Thus, future research should look into disentangling management quality and ESG performance by adding firm fixed effects, for instance. Of course, though, this would require much longer time series than are currently commercially available.

Finally, we apply our noise-correction procedure to aggregate ESG scores. It would be interesting to perform our estimation for E, S, and G scores separately, as well as for underlying fundamental ESG attributes, such as carbon emissions, workplace diversity, etc. Increasing the granularity of the scores would reduce the estimation error introduced by the aggregation implied in the ratings. It would also give one a sense of which pillar of ESG scores (E, S, or G) contains the largest amount of noise. It may also further inform the discussion on the materiality of certain indicators (see Khan, Serafeim, and Yoon, 2016) and the need for harmonized reporting standards.

8 Conclusions

It is notoriously difficult to measure the ESG performance of firms. ESG rating agencies often report different estimates for the same attribute. In this paper, we argue that a high level of noise in the estimates leads to a significant bias in the standard regressions that analyze the effects of ESG performance. An important institutional feature of the market for ESG ratings is that there are numerous raters, who use different inputs and methodologies in computing their ratings. We exploit this feature and propose an instrumental variable approach to correcting the bias, which is predicated on using scores of different agencies as noisy measures of true ESG performance.

We show that standard regression estimates of the effects of ESG on stock returns are downward biased and, on average, more than double once we apply our noise-correction procedure. We run our estimation separately for all raters in our sample, across four geographical regions, and in the majority of these regressions we observe an increase in the estimates. Importantly, this result is stable across multiple horizons over which we measure returns.

The practical takeaway of these results is that it is worthwhile to rely on several complementary ESG ratings. While we find the scores of some rating agencies to be very noisy, it does not mean that they are uninformative. Two illustrative examples are Truvalue Labs and S&P Global. Our estimation procedure shows that while these scores do not perform well as predictors of stock returns, they are nevertheless valuable instruments that enhance the prediction of other scores.

We provide a ranking of ESG rating agencies' scores, from the least noisy to the noisiest. One may be tempted to conclude from our results that one should use ESG scores containing the least amount of noise, rather than scores of other raters. We caution against such an interpretation for two reasons. First, this ranking is specific to our model and regression setup and should not be overgeneralized. Second, our results show that for the least noisy ratings, coefficients increase substantially when instrumented with other ratings. In other words, relying on the scores of several complementary ratings yields better results.

Our paper offers a practical solution to deal with the divergence of ESG ratings. Whenever an ESG rating is used as a regressor, attenuation bias is likely to become a problem. If a second ESG rating is available, it can be used as an instrument, which reduces the attenuation bias. In this case, one needs to defend the assumption that the measurement error of the other ESG rating is orthogonal. If more than one additional ESG rating is available, one can rely on the OIR test to check the validity of instruments. This 2SLS approach to noise reduction is superior to using the averages of ESG scores or principal component analysis. And if noise is indeed a problem, it will make the empirical results stronger.

References

Adams, Renée B., and Daniel Ferreira, 2009, Women in the boardroom and their impact on governance and performance, *Journal of Financial Economics* 94, 291–309.

Albuquerque, Rui, Yrjö Koskinen, and Chendi Zhang, 2019, Corporate social responsibility and firm risk: Theory and empirical evidence, *Management Science* 65, 4451–4469.

Atz, Ulrich, Christopher Bruno, Zongyuan Zoe Liu, and Tracy Van Holt, 2022, Does sustainability generate better financial performance? Review, meta-analysis, and propositions, *Journal of Sustainable Finance and Investment (Forthcoming)*.

Avramov, Doron, Si Cheng, Abraham Lioui, and Andrea Tarelli, 2021, Sustainable investing with ESG rating uncertainty, *Journal of Financial Economics (Forthcoming)*.

Bauer, Rob, Nadja Guenster, and Rogér Otten, 2004, Empirical evidence on corporate governance in Europe: The effect on stock returns, firm value and performance, *Journal of Asset Management* 5, 91–104.

Bebchuk, Lucian, Alma Cohen, and Allen Ferrell, 2009, What matters in corporate governance?, *The Review of Financial Studies* 22, 783–827.

Berg, Florian, Kornelia Fabisik, and Zacharias Sautner, 2021, Is history repeating itself? The (un)predictable past of ESG ratings, *Working Paper, ECGI* Available at: https://ssrn.com/abstract=3722087.

Berg, Florian, Julian F. Kölbel, and Roberto Rigobon, 2020, Aggregate confusion: The divergence of ESG ratings, *Working Paper, MIT Sloan* Available at: https://papers.ssrn.com/abstract=3438533.

Bolton, Patrick, and Marcin T. Kacperczyk, 2020, Carbon premium around the world, *Working Paper, Imperial College* Available at: https://www.ssrn.com/abstract=3550233.

Broccardo, Eleonora, Oliver Hart, and Luigi Zingales, 2022, Exit vs. voice, *Journal of Political Economy (Forthcoming)*.

Chava, Sudheer, 2014, Environmental externalities and cost of capital, *Management Science* 60, 2223–2247.

Christensen, Dane M., George Serafeim, and Anywhere Sikochi, 2022, Why is corporate virtue in the eye of the beholder? The case of ESG ratings, *The Accounting Review* 97, 147–175.

Christensen, Hans Bonde, Luzi Hail, and Christian Leuz, 2021, Mandatory CSR and sustainability reporting: Economic analysis and literature review, *Working Paper, University of Chicago* Available at: https://papers.ssrn.com/abstract=3427748.

Daines, Robert M., Ian D. Gow, and David F. Larcker, 2010, Rating the ratings: How good are commercial governance ratings?, *Journal of Financial Economics* 98, 439–461.

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

El Ghoul, Sadok, Omrane Guedhami, Chuck C.Y. Kwok, and Dev R. Mishra, 2011, Does corporate social responsibility affect the cost of capital?, *Journal of Banking & Finance* 35, 2388–2406.

Fama, Eugene F., and Kenneth R. French, 2007, Disagreement, tastes, and asset prices, *Journal of Financial Economics* 83, 667–689.

Friedman, Henry L., and Mirko S. Heinle, 2016, Taste, information, and asset prices: Implications for the valuation of CSR, *Review of Accounting Studies* 22, 740–767.

Gibson, Rajna, Philipp Krueger, and Peter Steffen Schmidt, 2021, ESG rating disagreement and stock returns, *Financial Analyst Journal (forthcoming).*

Gompers, Paul A., Joy L. Ishii, and Andrew Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics* 118, 107–155.

Heinkel, Robert, Alan Kraus, and Josef Zechner, 2001, The effect of green investment on corporate behavior, *Journal of Financial and Quantitative Analysis* 36, 431–449.

Kashyap, Anil K., Natalia Kovrijnykh, Jian Li, and Anna Pavlova, 2021, The benchmark inclusion subsidy, *Journal of Financial Economics* 142, 756–774.

Khan, Mozaffar, George Serafeim, and Aaron Yoon, 2016, Corporate sustainability: First evidence on materiality, *The Accounting Review* 91, 1697–1724.

Landier, Augustin, and Stefano Lovo, 2020, ESG investing: How to optimize impact?, *Working Paper, HEC Paris* Available at: https://www.ssrn.com/abstract=3508938.

Larcker, David F., Peter C. Reiss, and Youfei Xiao, 2015, Corporate governance data and measures revisited, *Working Paper, Stanford University* Available at: https://papers.ssrn.com/abstract=2694802.

Larcker, David F., Scott A. Richardson, and Irem Tuna, 2007, Corporate governance, accounting outcomes, and organizational performance, *The Accounting Review* 82, 963–1008.

Lewellen, Jonathan, 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Lins, Karl V., Henri Servaes, and Ane M. Tamayo, 2017, Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis, *Journal of Finance* 27, 1785–1824.

McCahery, Joseph A., Zacharias Sautner, and Laura T. Starks, 2016, Behind the scenes: The corporate governance preferences of institutional investors, *Journal of Finance* 71, 2905–2932.

Oehmke, Martin, and Marcus M. Opp, 2019, A theory of socially responsible investment, *Working Paper, Swedish House of Finance* Available at: https://www.ssrn.com/abstract=3467644.

Pancost, N. Aaron, and Garrett Schaller, 2021, Measuring measurement error, *Working Paper* Available at: https://www.ssrn.com/abstract=4045772.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2021a, Dissecting green returns, *Working Paper, National Bureau of Economic Research*.

——— , 2021b, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, 2021, Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics* 142, 572–597.

Strike, Vanessa M., Jijun Gao, and Pratima Bansal, 2006, Being good while being bad: Social responsibility and the international diversification of US firms, *Journal of International Business Studies* 37, 850–862.

A Internet Appendix

A.1 Derivation of Estimates

A.1.1 OLS and IV

Under the assumption that the controls are orthogonal to all RHS variables, the OLS estimate of β in Equation (3) is given by

$$\beta_{OLS} = \frac{s'_{k,t,i} r_{k,t+1}}{s'_{k,t,i} s_{k,t,i}} = \frac{cov(r_{k,t+1}, s_{k,t,i})}{var(s_{k,t,i})}$$

Using the structural Equations (1) and (2), the covariance between the stock returns and the ESG score is

$$\begin{aligned} cov(r_{k,t+1}, s_{k,t,i}) &= cov(\beta Y_{k,t} + M_{k,t}, Y_{k,t}) \\ &= \beta var(Y_{k,t}) + cov(M_{k,t}, Y_{k,t}) \\ &= \beta var(Y_{k,t}) + \gamma var(M_{k,t}) \end{aligned}$$

The variance of the score is given by

$$\begin{aligned} var(s_{k,t,i}) &= var(Y_{k,t} + \eta_{k,t,i}) \\ &= var(Y_{k,t}) + var(\eta_{k,t,i}) \end{aligned}$$

The ratio of these two quantities is the OLS estimate:

$$\begin{aligned} \beta_{OLS} &= \frac{\beta \cdot var(Y_{k,t}) + \gamma \cdot var(M_{k,t})}{var(Y_{k,t}) + var(\eta_{k,t,i})} \\ &= \underbrace{\frac{var(Y_{k,t})}{var(Y_{k,t}) + var(\eta_{k,t,i})}}_{\text{Attenuation}} \cdot \left[\beta + \underbrace{\gamma \frac{var(M_{k,t})}{var(Y_{k,t})}}_{\text{Omitted Variable}} \right] \end{aligned}$$

The derivation of IV estimate is as follows: The structural equations of the instrumental variable estimate are

$$\begin{aligned}
\beta_{IV} &= \frac{cov(r_{k,t+1}, s_{k,t,j})}{cov(s_{k,t,i},, s_{k,t,j})} \\
r_{k,t+1} &= \alpha + \beta \cdot Y_{k,t} + M_{k,t} + \epsilon_{k,t} \\
s_{k,t,i} &= Y_{k,t} + \eta_{k,t,i} \\
s_{k,t,j} &= Y_{k,t} + \eta_{k,t,j}
\end{aligned}$$

where the score identified from rating agency i is the regressor, while the score from the rating agency $j \neq i$ is the instrument. The covariance between the stock returns and the ESG score used for the instrumentation ($s_{k,t,j}$) when the measurement errors are orthogonal to all controls and innovations is as follows:

$$\begin{aligned}
cov(r_{k,t+1}, s_{k,t,j}) =& cov(\beta Y_{k,t} + M_{k,t} + \epsilon_{k,t}, Y_{k,t} + \eta_{k,t,j}) \\
=& \beta var(Y_{k,t}) + cov(M_{k,t}, Y_{k,t}) \\
=& \beta var(Y_{k,t}) + \gamma var(M_{k,t})
\end{aligned}$$

The covariance between the ESG score from rating i and the ESG score used for the instrumentation ($s_{k,t,j}$) is as follows:

$$\begin{aligned}
cov(s_{k,t,i}, s_{k,t,j}) &= cov(Y_{k,t} + \eta_{k,t,i}, Y_{k,t} + \eta_{k,t,j}) \\
&= var(Y_{k,t})
\end{aligned}$$

The IV estimate is then

$$\begin{aligned}
\beta_{IV} =& \frac{\beta var(Y_{k,t}) + \gamma var(M_{k,t})}{var(Y_{k,t})} \\
=& \beta + \gamma \frac{var(M_{k,t})}{var(Y_{k,t})}
\end{aligned}$$

A.1.2 Biased and Inconsistent Estimates of OLS and IV

Derivation of OLS when identifying restrictions are violated. In all these derivations we drop the constants and the controls to simplify the notation. Including them in the derivation does not change the results. All variables are assumed to have mean zero (hence, the constants are not needed). Under the assumption that the controls are orthogonal to all RHS variables, the OLS estimate of β in Equation (3) is given by

$$\begin{aligned}
\beta_{OLS} &= \frac{s'_{k,t,i} r_{k,t+1}}{s'_{k,t,i} s_{k,t,i}} = \frac{cov(r_{k,t+1}, s_{k,t,i})}{var(s_{k,t,i})} \\
r_{k,t+1} &= \alpha + \beta \cdot Y_{k,t} + M_{k,t} + \epsilon_{k,t} \\
s_{k,t,i} &= Y_{k,t} + \eta_{k,t,i}
\end{aligned}$$

The covariance between the stock returns and the ESG score is

$$\begin{aligned}
cov(r_{k,t+1}, s_{k,t,i}) &= cov(\beta Y_{k,t} + M_{k,t} + \epsilon_{k,t}, Y_{k,t} + \eta_{k,t,i}) \\
&= \beta var(Y_{k,t}) + cov(M_{k,t}, Y_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,i}) + cov(M_{k,t}, \eta_{k,t,i}) + cov(\epsilon_{k,t}, \eta_{k,t,i}) \\
&= \beta var(Y_{k,t}) + \gamma var(M_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,i}) + cov(M_{k,t}, \eta_{k,t,i}) + cov(\epsilon_{k,t}, \eta_{k,t,i})
\end{aligned}$$

The variance of the score is given by

$$\begin{aligned}
var(s_{k,t,i}) &= var(Y_{k,t} + \eta_{k,t,i}) \\
&= var(Y_{k,t}) + var(\eta_{k,t,i}) + cov(Y_{k,t}, \eta_{k,t,i})
\end{aligned}$$

The ratio of these two quantities is the OLS estimate:

$$\beta_{OLS} = \frac{\beta var(Y_{k,t}) + \gamma var(M_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,i}) + cov(M_{k,t}, \eta_{k,t,i}) + cov(\epsilon_{k,t}, \eta_{k,t,i})}{var(Y_{k,t}) + var(\eta_{k,t,i}) + cov(Y_{k,t}, \eta_{k,t,i})}$$

The structural equations of the instrumental variable estimate are

$$\begin{aligned}
\beta_{IV} &= \frac{cov(r_{k,t+1}, s_{k,t,j})}{cov(s_{k,t,i},, s_{k,t,j})} \\
r_{k,t+1} &= \alpha + \beta \cdot Y_{k,t} + M_{k,t} + \epsilon_{k,t} \\
s_{k,t,i} &= Y_{k,t} + \eta_{k,t,i} \\
s_{k,t,j} &= Y_{k,t} + \eta_{k,t,j}
\end{aligned}$$

where the score identified from rating agency i is the regressor, while the score from the rating agency $j \neq i$ is the instrument. The covariance between the stock returns and the

ESG score used for the instrumentation ($s_{k,t,j}$) is as follows:

$$\begin{aligned} cov(r_{k,t+1}, s_{k,t,j}) =& cov(\beta Y_{k,t} + M_{k,t} + \epsilon_{k,t}, Y_{k,t} + \eta_{k,t,j}) \\ =& \beta var(Y_{k,t}) + cov(M_{k,t}, Y_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,j}) + \\ & cov(M_{k,t}, \eta_{k,t,j}) + cov(\epsilon_{k,t}, \eta_{k,t,j}) \\ =& \beta var(Y_{k,t}) + \gamma var(M_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,j}) + cov(M_{k,t}, \eta_{k,t,j}) + cov(\epsilon_{k,t}, \eta_{k,t,j}) \end{aligned}$$

The covariance between the ESG score from rating i and the ESG score used for the instrumentation ($s_{k,t,j}$) is the following

$$\begin{aligned} cov(s_{k,t,i}, s_{k,t,j}) &= cov(Y_{k,t} + \eta_{k,t,i}, (Y_{k,t} + \eta_{k,t,j}) \\ &= var(Y_{k,t}) + cov(Y_{k,t}, \eta_{k,t,j}) + cov(Y_{k,t}, \eta_{k,t,i}) + cov(\eta k, t, i, \eta_{k,t,j}) \end{aligned}$$

The IV estimate is then

$$\beta_{IV} = \frac{\beta var(Y_{k,t}) + \gamma var(M_{k,t}) + \beta cov(Y_{k,t}, \eta_{k,t,j}) + cov(M_{k,t}, \eta_{k,t,j}) + cov(\epsilon_{k,t}, \eta_{k,t,j})}{var(Y_{k,t}) + cov(Y_{k,t}, \eta_{k,t,j}) + cov(Y_{k,t}, \eta_{k,t,i}) + cov(\eta_{k,t,i}, \eta_{k,t,j})}$$

A.2 Proofs

Proof of Lemma 1. We start by showing (16)-(17). The conditional distribution of D given s_D is normal. The mean and the variance of that distribution can be computed by a linear regression of D on s_D:

$$D - \overline{D} = \beta_D(s_D - \overline{D}) + \varepsilon_D, \tag{39}$$

where $\varepsilon_D \sim N(0, \sigma^2_{\varepsilon_D})$ and is independent of s_D. We need to determine β_D.

The mean and variance of D conditional on signal s_D are

$$E(D|s_D) = \overline{D} + \beta_D(s_D - \overline{D}) \tag{40}$$

$$Var(D|s_D) = \sigma^2_{\varepsilon_D} \tag{41}$$

The regression coefficient β_D is given by the following standard expression:

$$\beta_D = \frac{Cov(D - \overline{D}, s_D - \overline{D})}{Var(s_D - \overline{D})} = \frac{(Cov(D - \overline{D}, D - \overline{D} + \eta_D)}{Var(D - \overline{D} + \eta_D)} = \frac{\sigma^2_D}{\sigma^2_D + \sigma^2_{\eta_D}} \tag{42}$$

The variables D and η_D are independent. Taking variances on each side of (39), we have

$$Var(D-\overline{D}) = Var(\beta_D(s_D-\overline{D})+\varepsilon_D) = \beta_D^2 Var(s_D-\overline{D})+\sigma_{\varepsilon_D}^2 \tag{43}$$

It is easy to see that

$$\sigma_{\varepsilon_D}^2 = \sigma_D^2 - \left(\frac{\sigma_D^2}{\sigma_D^2+\sigma_{\eta_D}^2}\right)^2(\sigma_D^2+\sigma_{\eta_D}^2) = \frac{\sigma_D^2\sigma_{\eta_D}^2}{\sigma_D^2+\sigma_{\eta_D}^2} \tag{44}$$

Hence, the mean and variance of D, conditional on signal s_D, are

$$E(D|s_D) = \overline{D}+\beta(s_D-\overline{D}) = \overline{D}+\frac{\sigma_D^2}{\sigma_D^2+\sigma_{\eta_D}^2}(s_D-\overline{D}) \tag{45}$$

$$Var(D|s_D) = \sigma_{\varepsilon_D}^2 = \frac{\sigma_D^2\sigma_{\eta_D}^2}{\sigma_D^2+\sigma_{\eta_D}^2} \tag{46}$$

The derivation for the mean and variance of Y, conditional on signal s_Y, is analogous. In the equations above, we need to replace D with Y and s_D with s_Y. □

Proof of Lemma 2. An ESG-conscious investor chooses their portfolio θ^{ESG} to maximize the expected utility

$$E\left(-\exp(-\gamma(W_1+\theta^{ESG}Y)\,|s_D,s_Y\right).$$

Substituting in their wealth in period 1, we arrive at

$$E\left(-\exp(-\gamma(W_0+\theta^{ESG}(D-p)+\theta^{ESG}Y)\,|s_D,s_Y\right).$$

For a normally distributed random variable x, $E(\exp(x)) = \exp\left(E(x)+\frac{1}{2}Var(x)\right)$. Since D and Y are independent, normally distributed random variables, we can show that the above objective is equivalent to the following mean-variance optimization:

$$\max_{\theta^{ESG}} \theta^{ESG}\left[(E(D\,|s_D,s_Y)-p)+E(Y\,|s_D,s_Y)\right]-\frac{1}{2}\gamma(\theta^{ESG})^2\left[Var(D\,|s_D,s_Y)+Var(Y\,|s_D,s_Y)\right].$$

Solving for the portfolio choice θ^{ESG} that maximizes the above objective, we arrive at (21).

To solve for the portfolio of traditional investors, we simply repeat the above derivations, setting Y equal to zero. □

Proof of Proposition 1. Substituting in θ^{ESG} and θ^{T} from Lemma 2 into market clearing (22), we derive

$$\begin{aligned} p_0 =& A\lambda Var(D|s_D)(E(D|s_D) + E(Y|s_Y)) + A(1-\lambda)(Var(D|s_D) + Var(Y|s_Y))E(D|s_D) \\ & - A\gamma\overline{\theta} Var(D|s_D)(Var(D|s_D) + Var(Y|s_Y)) \end{aligned}$$

where

$$A = \left[\lambda Var(D|s_D) + (1-\lambda)(Var(D|s_D) + Var(Y|s_Y))\right]^{-1}.$$

Substituting the expressions for the conditional moments from Lemmma 1, we have

$$\begin{aligned} p_0 =& A\lambda \frac{\sigma_D^2 \sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} \left(\overline{Y} + \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_{\eta Y}^2}(s_Y - \overline{Y}) + \overline{D} + \frac{\sigma_D^2}{\sigma_D^2 + \sigma_{\eta D}^2}(s_D - \overline{D}) \right) \\ & + A(1-\lambda) \left(\frac{\sigma_D^2 \sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} + \frac{\sigma_Y^2 \sigma_{\eta Y}^2}{\sigma_Y^2 + \sigma_{\eta Y}^2} \right) \left(\overline{D} + \frac{\sigma_D^2}{\sigma_D^2 + \sigma_{\eta D}^2}(s_D - \overline{D}) \right) \\ & - A\gamma\overline{\theta} \frac{\sigma_D^2 \sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} \left[\frac{\sigma_D^2 \sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} + \frac{\sigma_Y^2 \sigma_{\eta Y}^2}{\sigma_Y^2 + \sigma_{\eta Y}^2} \right], \end{aligned}$$

where

$$A = \left[\frac{\sigma_D^2 \sigma_{\eta D}^2}{\sigma_D^2 + \sigma_{\eta D}^2} + (1-\lambda) \frac{\sigma_Y^2 \sigma_{\eta Y}^2}{\sigma_Y^2 + \sigma_{\eta Y}^2} \right]^{-1}.$$

Simplifying the above expression, we arrive at the statement in the proposition. □

A.3 Additional Tables from the Estimation

Table A1. OLS estimates for 1- to 8-month returns. This table reports estimates of β from the OLS regression (27), with $h = 1, \ldots, 8$.. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		OLS																							
		1 month			2 months			3 months			4 months			5 months			6 months			7 months			8 months		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.152	0.071	**	0.155	0.072	**	0.159	0.073	**	0.163	0.073	**	0.171	0.074	**	0.180	0.076	**	0.185	0.076	**	0.187	0.076	**
	MSCI	0.123	0.053	**	0.135	0.051	***	0.145	0.052	***	0.152	0.053	***	0.154	0.054	***	0.159	0.055	***	0.163	0.056	***	0.165	0.057	***
	Refinitiv	0.012	0.066		0.009	0.065		0.002	0.065		0.006	0.065		0.007	0.066		0.011	0.067		0.018	0.069		0.019	0.070	
	RepRisk	0.146	0.069	**	0.145	0.068	**	0.142	0.068	**	0.128	0.068	*	0.118	0.068	*	0.111	0.068		0.097	0.069		0.092	0.069	
	S&P Global	0.122	0.083		0.116	0.083		0.111	0.084		0.115	0.083		0.116	0.082		0.115	0.081		0.117	0.081		0.111	0.081	
	Sustainalytics	0.123	0.076		0.125	0.077		0.132	0.076	*	0.135	0.076	*	0.131	0.077	*	0.128	0.077		0.125	0.078		0.124	0.078	
	TVL	0.039	0.066		0.036	0.066		0.041	0.065		0.041	0.063		0.039	0.062		0.041	0.060		0.038	0.059		0.038	0.059	
	Moody's	0.027	0.076		0.028	0.078		0.031	0.079		0.036	0.079		0.042	0.079		0.047	0.079		0.054	0.080		0.056	0.080	
GBP	ISS	-0.265	0.132	**	-0.261	0.129	**	-0.271	0.127	**	-0.285	0.127	**	-0.296	0.131	**	-0.280	0.132	**	-0.267	0.131	**	-0.264	0.129	**
	MSCI	-0.153	0.100		-0.170	0.098	*	-0.178	0.099	*	-0.186	0.101	*	-0.174	0.103		-0.158	0.102		-0.148	0.100		-0.142	0.100	
	Refinitiv	-0.122	0.109		-0.120	0.105		-0.151	0.108		-0.177	0.111		-0.191	0.112	*	-0.186	0.115		-0.179	0.117		-0.188	0.117	
	RepRisk	0.394	0.090	***	0.398	0.090	***	0.391	0.084	***	0.377	0.081	***	0.371	0.079	***	0.382	0.080	***	0.378	0.079	***	0.379	0.078	***
	S&P Global	-0.035	0.115		-0.058	0.115		-0.093	0.115		-0.119	0.117		-0.116	0.119		-0.112	0.119		-0.099	0.120		-0.106	0.121	
	Sustainalytics	-0.025	0.128		-0.003	0.127		-0.020	0.127		-0.041	0.128		-0.061	0.133		-0.086	0.137		-0.088	0.138		-0.088	0.139	
	TVL	0.087	0.078		0.093	0.074		0.093	0.076		0.105	0.076		0.129	0.077		0.145	0.079	*	0.153	0.080	*	0.163	0.081	**
	Moody's	-0.142	0.122		-0.157	0.122		-0.188	0.129		-0.216	0.136		-0.223	0.138		-0.216	0.140		-0.207	0.140		-0.221	0.141	
JPY	ISS	0.134	0.071	*	0.133	0.070	*	0.122	0.069	*	0.119	0.070	*	0.125	0.070	*	0.129	0.070	*	0.142	0.070	**	0.149	0.070	**
	MSCI	0.092	0.057		0.084	0.057		0.073	0.056		0.073	0.055		0.080	0.054		0.080	0.054		0.085	0.053		0.090	0.053	*
	Refinitiv	0.187	0.057	***	0.187	0.055	***	0.181	0.054	***	0.179	0.053	***	0.180	0.052	***	0.185	0.051	***	0.190	0.051	***	0.195	0.050	***
	RepRisk	0.096	0.063		0.089	0.060		0.085	0.061		0.085	0.061		0.092	0.061		0.092	0.060		0.095	0.059		0.096	0.059	
	S&P Global	0.093	0.074		0.117	0.075		0.114	0.075		0.113	0.075		0.114	0.074		0.119	0.075		0.121	0.075		0.124	0.075	
	Sustainalytics	0.258	0.075	***	0.261	0.070	***	0.261	0.068	***	0.259	0.066	***	0.258	0.065	***	0.260	0.065	***	0.264	0.065	***	0.266	0.064	***
	TVL	0.033	0.054		0.038	0.053		0.040	0.053		0.038	0.053		0.035	0.052		0.037	0.052		0.036	0.051		0.037	0.050	
	Moody's	0.047	0.060		0.060	0.060		0.067	0.059		0.066	0.058		0.067	0.058		0.081	0.057		0.093	0.058		0.104	0.059	*
USD	ISS	0.069	0.055		0.082	0.057		0.082	0.057		0.083	0.057		0.083	0.057		0.085	0.057		0.084	0.058		0.086	0.057	
	MSCI	0.067	0.041		0.067	0.042		0.069	0.041		0.069	0.041	*	0.071	0.041	*	0.069	0.042		0.067	0.043		0.068	0.043	
	Refinitiv	0.064	0.055		0.066	0.054		0.060	0.054		0.057	0.054		0.056	0.054		0.054	0.054		0.053	0.054		0.054	0.054	
	RepRisk	0.022	0.056		0.053	0.057		0.066	0.056		0.075	0.055		0.083	0.054		0.091	0.053	*	0.093	0.052	*	0.094	0.052	*
	S&P Global	0.012	0.045		0.011	0.045		0.011	0.046		0.015	0.046		0.018	0.046		0.019	0.045		0.017	0.045		0.016	0.045	
	Sustainalytics	0.140	0.069	**	0.140	0.072	*	0.135	0.073	*	0.139	0.074	*	0.145	0.074	*	0.149	0.074	**	0.150	0.073	**	0.145	0.073	**
	TVL	0.051	0.046		0.052	0.046		0.051	0.046		0.046	0.045		0.046	0.045		0.039	0.045		0.033	0.044		0.028	0.044	
	Moody's	-0.027	0.056		-0.018	0.056		-0.019	0.055		-0.014	0.055		-0.010	0.054		-0.008	0.055		-0.007	0.054		-0.006	0.054	

Table A2. Pruned IV estimates for 1- to 8-month returns. This table reports estimates of β from 2SLS regression (29), with $h = 1, \ldots, 8$.. The Pruning procedure is described in Section 3.3. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

		IV Pruned																							
		1 month			2 months			3 months			4 months			5 months			6 months			7 months			8 months		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.137	0.054	***	0.341	0.051	***	0.079	0.067		0.052	0.065		0.058	0.066		0.066	0.068		0.081	0.070		0.083	0.071	
	MSCI	0.194	0.071	***	0.332	0.070	***	0.339	0.071	***	0.330	0.071	***	0.330	0.072	***	0.310	0.070	***	0.304	0.070	***	0.295	0.069	***
	Refinitiv	0.183	0.071	***	0.304	0.071	***	0.310	0.072	***	0.321	0.072	***	0.326	0.073	***	0.314	0.074	***	0.317	0.075	***	0.266	0.080	***
	RepRisk	0.333	0.072	***	0.446	0.052	***	0.398	0.062	***	0.399	0.062	***	0.376	0.062	***	0.114	0.066	*	0.093	0.068		-	-	
	S&P Global	0.055	0.055		0.065	0.066		0.034	0.066		0.043	0.065		0.050	0.066		0.059	0.068		0.075	0.070		0.077	0.071	
	Sustainalytics	0.237	0.071	***	0.422	0.071	***	0.429	0.072	***	0.432	0.072	***	0.410	0.054	***	0.411	0.055	***	0.408	0.056	***	0.354	0.069	***
	TVL	0.634	0.071	***	0.753	0.071	***	0.759	0.072	***	0.811	0.071	***	0.472	0.066	***	0.459	0.067	***	0.425	0.069	***	0.409	0.070	***
	Moody's	0.220	0.071	***	0.269	0.072	***	0.311	0.072	***	0.321	0.072	***	0.325	0.073	***	0.332	0.075	***	0.318	0.075	***	0.387	0.076	***
GBP	ISS	-0.218	0.101	**	-0.231	0.099	**	-0.281	0.099	***	-0.293	0.113	***	-0.325	0.102	***	-0.320	0.100	***	-0.305	0.098	***	-0.318	0.098	***
	MSCI	-0.318	0.136	**	-0.117	0.107		-0.182	0.110		-0.288	0.115	**	-0.345	0.117	***	-0.290	0.120	**	-0.287	0.121	**	-0.293	0.123	**
	Refinitiv	-0.309	0.132	**	-0.329	0.129	***	-0.455	0.126	***	-0.442	0.125	***	-0.540	0.130	***	-0.509	0.130	***	-0.483	0.129	***	-0.487	0.127	***
	RepRisk	0.194	0.138		0.163	0.103		0.494	0.110	***	0.552	0.114	***	0.527	0.115	***	0.487	0.120	***	0.492	0.120	***	0.526	0.122	***
	S&P Global	-0.313	0.133	**	-0.315	0.129	**	-0.385	0.126	***	-0.436	0.127	***	-0.463	0.130	***	-0.415	0.131	***	-0.397	0.130	***	-0.410	0.128	***
	Sustainalytics	-0.457	0.133	***	-0.487	0.129	***	-0.525	0.126	***	-0.611	0.126	***	-0.603	0.130	***	-0.651	0.131	***	-0.614	0.130	***	-0.610	0.128	***
	TVL	0.705	0.127	***	0.916	0.121	***	0.973	0.120	***	0.931	0.083	***	0.901	0.081	***	0.973	0.080	***	0.960	0.079	***	0.964	0.078	***
	Moody's	-0.252	0.133	*	-0.265	0.130	**	-0.396	0.128	***	-0.486	0.128	***	-0.442	0.133	***	-0.431	0.133	***	-0.412	0.132	***	-0.412	0.130	***
JPY	ISS	0.143	0.056	***	0.168	0.056	***	0.170	0.054	***	0.158	0.054	***	0.161	0.053	***	0.179	0.053	***	0.195	0.052	***	0.210	0.051	***
	MSCI	0.477	0.068	***	0.491	0.067	***	0.409	0.067	***	0.404	0.068	***	0.362	0.068	***	0.376	0.068	***	0.393	0.068	***	0.406	0.068	***
	Refinitiv	0.104	0.071		0.220	0.068	***	0.160	0.067	**	0.248	0.068	***	0.254	0.067	***	0.263	0.067	***	0.274	0.067	***	0.193	0.051	***
	RepRisk	0.063	0.074		-0.061	0.059		0.175	0.057	***	0.165	0.056	***	0.058	0.052		-	-		-	-		-	-	
	S&P Global	0.330	0.068	***	0.325	0.068	***	0.304	0.067	***	0.299	0.068	***	0.308	0.068	***	0.316	0.068	***	0.332	0.068	***	0.320	0.069	***
	Sustainalytics	0.473	0.069	***	0.492	0.067	***	0.468	0.066	***	0.399	0.068	***	0.413	0.068	***	0.356	0.052	***	0.376	0.051	***	0.395	0.050	***
	TVL	0.707	0.072	***	0.681	0.071	***	0.646	0.070	***	0.533	0.072	***	0.465	0.053	***	0.478	0.053	***	0.966	0.053	***	1.047	0.052	***
	Moody's	0.265	0.068	***	0.250	0.067	***	0.237	0.066	***	0.227	0.067	***	0.232	0.067	***	0.238	0.066	***	0.248	0.066	***	0.288	0.068	***
USD	ISS	0.021	0.041		0.093	0.041	**	-0.001	0.056		0.007	0.055		0.073	0.053		0.071	0.053		0.011	0.046		0.010	0.045	
	MSCI	0.149	0.057	***	0.202	0.058	***	0.214	0.059	***	0.153	0.059	***	0.155	0.058	***	0.150	0.059	***	0.141	0.059	**	0.167	0.059	***
	Refinitiv	0.032	0.057		-0.004	0.041		-0.018	0.056		0.003	0.046		0.009	0.046		0.012	0.046		0.010	0.046		0.009	0.045	
	RepRisk	0.036	0.057		-0.049	0.059		0.201	0.046	***	0.160	0.046	***	0.142	0.046	***	0.105	0.046	**	0.088	0.046	*	0.067	0.045	
	S&P Global	0.083	0.057		0.139	0.058	**	0.117	0.058	**	0.124	0.059	**	0.124	0.058	**	0.124	0.059	**	0.122	0.059	**	0.125	0.059	**
	Sustainalytics	0.168	0.057	***	0.256	0.059	***	0.353	0.060	***	0.354	0.060	***	0.324	0.060	***	0.247	0.042	***	0.238	0.042	***	0.312	0.060	***
	TVL	0.294	0.057	***	0.367	0.058	***	0.367	0.059	***	0.381	0.059	***	0.399	0.059	***	0.405	0.059	***	0.404	0.059	***	0.412	0.059	***
	Moody's	0.098	0.057	*	0.097	0.058		0.113	0.058	*	0.096	0.059		0.114	0.058	**	0.114	0.059	*	0.112	0.059	*	0.115	0.059	*

Table A3. Lasso IV estimates for for 1- to 8-month returns. This table reports estimates of β from the 2SLS regression (29), with $h = 1, \ldots, 8$.. The Lasso procedure is described in Section 3.3. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		IV Lasso																							
		1 month			2 months			3 months			4 months			5 months			6 months			7 months			8 months		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.137	0.054	***	0.076	0.067		0.071	0.066		0.080	0.066		0.050	0.066		0.059	0.068		0.073	0.070		0.076	0.071	
	MSCI	0.194	0.071	***	0.114	0.067	*	0.117	0.067	*	0.045	0.065		0.052	0.066		0.060	0.068		0.075	0.070		0.078	0.071	
	Refinitiv	0.183	0.071	***	0.235	0.072	***	-	-		-	-		-	-		-	-		0.155	0.080	*	0.151	0.080	*
	RepRisk	0.333	0.072	***	0.203	0.063	***	-0.061	0.065		-	-		-	-		-	-		-	-		-	-	
	S&P Global	0.103	0.072		0.039	0.066		0.035	0.066		0.045	0.065		0.052	0.066		0.061	0.068		0.076	0.070		0.079	0.071	
	Sustainalytics	0.237	0.071	***	0.239	0.052	***	-	-		-	-		-	-		-	-		-	-		-	-	
	TVL	0.634	0.071	***	0.648	0.072	***	0.803	0.071	***	0.809	0.071	***	0.722	0.054	***	-	-		-	-		-	-	
	Moody's	0.220	0.071	***	0.241	0.072	***	-	-		-	-		-	-		-	-		-	-		-	-	
GBP	ISS	-0.239	0.101	**	-0.231	0.099	**	-0.281	0.099	***	-0.330	0.100	***	-0.347	0.101	***	-0.320	0.100	***	-0.305	0.098	***	-0.318	0.098	***
	MSCI	0.017	0.141		-0.394	0.132	***	-0.432	0.129	***	-	-		-	-		-	-		-0.515	0.132	***	-	-	
	Refinitiv	-0.328	0.132	**	-0.325	0.129	***	-0.364	0.126	***	-0.396	0.126	***	-0.435	0.130	***	-0.416	0.130	***	-	-		-	-	
	RepRisk	0.194	0.138		0.201	0.134		0.193	0.104	*	-	-		-	-		-	-		-	-		-	-	
	S&P Global	-0.324	0.133	**	-0.315	0.129	**	-0.364	0.126	***	-0.414	0.127	***	-0.463	0.130	***	-0.456	0.130	***	-0.439	0.129	***	-0.454	0.127	***
	Sustainalytics	-0.340	0.135	***	-0.546	0.128	***	-0.596	0.126	***	-0.646	0.126	***	-0.644	0.130	***	-0.606	0.131	***	-0.568	0.130	***	-0.568	0.128	***
	TVL	0.705	0.127	***	-	-		-	-		-	-		-	-		-	-		-	-		-	-	
	Moody's	-0.330	0.131	***	-0.255	0.130	**	-0.297	0.127	**	-0.337	0.127	***	-0.333	0.132	***	-0.337	0.131	***	-	-		-	-	
JPY	ISS	0.187	0.055	***	0.143	0.076	*	0.149	0.076	*	0.148	0.076	*	0.149	0.076	**	0.169	0.076	**	0.185	0.076	**	0.200	0.076	***
	MSCI	0.293	0.068	***	-	-		-	-		-	-		-	-		-	-		-	-		-	-	
	Refinitiv	0.128	0.070	*	0.150	0.069	**	0.153	0.068	**	0.150	0.069	**	0.136	0.076	*	0.155	0.076	**	0.170	0.076	**	0.185	0.076	**
	RepRisk	-0.327	0.058	***	-0.170	0.079	**	-	-		-	-		-	-		-	-		-	-		-	-	
	S&P Global	0.173	0.070	**	-	-		-	-		-	-		-	-		-	-		-	-		-	-	
	Sustainalytics	0.333	0.070	***	0.335	0.055	***	0.326	0.054	***	0.323	0.053	***	0.334	0.052	***	0.355	0.052	***	0.376	0.051	***	0.395	0.050	***
	TVL	0.730	0.072	***	0.683	0.055	***	0.647	0.054	***	-	-		-	-		-	-		-	-		-	-	
	Moody's	0.265	0.068	***	0.243	0.067	***	0.231	0.066	***	0.227	0.067	***	0.231	0.067	***	0.238	0.066	***	0.247	0.066	***	0.255	0.067	***
USD	ISS	0.021	0.041		0.037	0.041		0.014	0.053		0.019	0.053		0.024	0.053		0.007	0.046		0.006	0.046		0.006	0.045	
	MSCI	0.149	0.057	***	0.059	0.059		-	-		-	-		-	-		-	-		-	-		-	-	
	Refinitiv	0.032	0.057		0.009	0.059		-0.013	0.046		-0.004	0.046		0.002	0.046		0.006	0.046		0.005	0.046		0.005	0.045	
	RepRisk	0.062	0.057		0.004	0.041		-0.048	0.053		-0.055	0.053		-0.058	0.053		-0.074	0.053		-0.082	0.053		-0.103	0.053	*
	S&P Global	0.083	0.057		0.052	0.058		-	-		-	-		-	-		-	-		-	-		-	-	
	Sustainalytics	0.168	0.057	***	0.185	0.059	***	0.193	0.041	***	0.167	0.040	***	0.181	0.041	***	0.231	0.041	***	0.222	0.042	***	0.220	0.043	***
	TVL	0.294	0.057	***	0.352	0.058	***	0.351	0.058	***	0.351	0.059	***	0.391	0.059	***	0.397	0.059	***	0.396	0.059	***	0.402	0.059	***
	Moody's	0.098	0.057	*	0.105	0.058	*	0.097	0.058		0.097	0.059		0.099	0.059		0.099	0.059		0.088	0.059		0.108	0.059	*

Table A4. OLS estimates for month-by-month returns. This table reports estimates of β from the OLS regression (27), with $r_{k,t+h}$ now measuring return in month $t+h$. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		OLS																							
		Month 1			Month 2			Month 3			Month 4			Month 5			Month 6			Month 7			Month 8		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.152	0.071	**	0.148	0.074	**	0.146	0.074	*	0.152	0.075	**	0.179	0.081	**	0.195	0.084	**	0.195	0.085	**	0.194	0.081	**
	MSCI	0.123	0.053	**	0.146	0.052	***	0.168	0.056	***	0.166	0.058	***	0.168	0.059	***	0.183	0.062	***	0.188	0.065	***	0.194	0.065	***
	Refinitiv	0.012	0.066		-0.002	0.066		-0.003	0.066		0.005	0.067		-0.002	0.071		0.007	0.078		0.032	0.080		0.004	0.076	
	RepRisk	0.146	0.069	**	0.144	0.069	**	0.131	0.070	*	0.094	0.073		0.097	0.073		0.114	0.074		0.083	0.077		0.110	0.072	
	SPGlobal	0.122	0.083		0.090	0.084		0.093	0.085		0.102	0.084		0.108	0.085		0.103	0.085		0.104	0.091		0.050	0.090	
	Sustainalytics	0.123	0.076		0.124	0.080		0.129	0.078		0.127	0.075	*	0.106	0.080		0.123	0.081		0.126	0.084		0.139	0.077	*
	TVL	0.039	0.066		0.036	0.067		0.046	0.065		0.046	0.061		0.048	0.062		0.067	0.061		0.056	0.062		0.065	0.065	
	Moody's	0.027	0.076		0.002	0.082		0.025	0.082		0.034	0.081		0.028	0.084		0.033	0.085		0.046	0.089		0.042	0.091	
GBP	ISS	-0.265	0.132	**	-0.238	0.129	*	-0.253	0.123	**	-0.276	0.137	**	-0.281	0.153	*	-0.203	0.142		-0.178	0.140		-0.191	0.134	
	MSCI	-0.153	0.100		-0.190	0.102	*	-0.168	0.109		-0.208	0.111	*	-0.152	0.114		-0.101	0.108		-0.120	0.112		-0.144	0.115	
	Refinitiv	-0.122	0.109		-0.105	0.108		-0.111	0.115		-0.102	0.126		-0.083	0.125		-0.021	0.138		0.001	0.133		-0.059	0.139	
	RepRisk	0.394	0.090	***	0.394	0.094	***	0.430	0.094	***	0.375	0.100	***	0.360	0.091	***	0.337	0.095	***	0.245	0.093	***	0.254	0.088	***
	SPGlobal	-0.035	0.115		-0.071	0.118		-0.090	0.117		-0.122	0.126		-0.031	0.129		-0.051	0.125		-0.013	0.127		-0.093	0.136	
	Sustainalytics	-0.025	0.128		0.008	0.126		-0.011	0.133		-0.018	0.133		-0.024	0.139		-0.071	0.137		-0.050	0.137		-0.063	0.138	
	TVL	0.087	0.078		0.090	0.073		0.110	0.078		0.149	0.076	**	0.193	0.087	**	0.202	0.090	**	0.186	0.095	*	0.216	0.092	**
	Moody's	-0.142	0.122		-0.147	0.125		-0.175	0.127		-0.194	0.136		-0.151	0.128		-0.131	0.119		-0.078	0.120		-0.147	0.124	
JPY	ISS	0.134	0.071	*	0.129	0.070	*	0.115	0.069		0.112	0.072		0.135	0.067	**	0.136	0.070	*	0.182	0.072	***	0.173	0.077	**
	MSCI	0.092	0.057		0.079	0.059		0.070	0.055		0.079	0.056		0.104	0.053	*	0.081	0.058		0.103	0.057	*	0.111	0.058	*
	Refinitiv	0.187	0.057	***	0.206	0.056	***	0.208	0.058	***	0.208	0.057	***	0.221	0.055	***	0.211	0.052	***	0.238	0.053	***	0.267	0.056	***
	RepRisk	0.096	0.063		0.086	0.060		0.089	0.066		0.101	0.066		0.123	0.063	*	0.114	0.062	*	0.131	0.062	**	0.117	0.070	
	SPGlobal	0.093	0.074		0.136	0.076	*	0.135	0.076	*	0.117	0.075		0.117	0.076		0.113	0.076		0.123	0.079		0.160	0.081	**
	Sustainalytics	0.258	0.075	***	0.257	0.072	***	0.252	0.075	***	0.250	0.072	***	0.241	0.070	***	0.251	0.071	***	0.267	0.068	***	0.266	0.068	***
	TVL	0.033	0.054		0.045	0.054		0.056	0.055		0.042	0.056		0.050	0.055		0.069	0.052		0.072	0.054		0.086	0.054	
	Moody's	0.047	0.060		0.084	0.063		0.101	0.062		0.078	0.062		0.084	0.061		0.123	0.059	**	0.154	0.060	***	0.182	0.067	***
USD	ISS	0.069	0.055		0.093	0.058		0.073	0.057		0.083	0.058		0.070	0.056		0.077	0.061		0.061	0.063		0.078	0.065	
	MSCI	0.067	0.041		0.060	0.043		0.056	0.042		0.050	0.044		0.056	0.046		0.048	0.048		0.052	0.048		0.073	0.050	
	Refinitiv	0.064	0.055		0.066	0.054		0.047	0.054		0.049	0.053		0.049	0.053		0.052	0.055		0.035	0.058		0.046	0.059	
	RepRisk	0.022	0.056		0.079	0.056		0.090	0.054		0.085	0.052		0.112	0.056	**	0.121	0.056	**	0.096	0.058		0.093	0.063	
	SPGlobal	0.012	0.045		-0.011	0.046		-0.017	0.046		-0.026	0.046		-0.034	0.046		-0.044	0.046		-0.061	0.048		-0.071	0.048	
	Sustainalytics	0.140	0.069	**	0.138	0.076	*	0.130	0.076	*	0.137	0.079	*	0.156	0.081	*	0.157	0.076	**	0.136	0.076	*	0.100	0.079	
	TVL	0.051	0.046		0.044	0.047		0.052	0.047		0.051	0.047		0.069	0.048		0.045	0.048		0.045	0.050		0.032	0.049	
	Moody's	-0.027	0.056		-0.014	0.054		-0.026	0.053		-0.010	0.052		-0.023	0.053		-0.020	0.054		-0.028	0.055		-0.028	0.056	

Table A5. Pruned IV estimates for month-by-month returns. This table reports estimates of β from 2SLS regression (29), with $r_{k,t+h}$ now measuring return in month $t+h$. The Pruning procedure is described in Section 3.3. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		IV Pruned																							
		Month 1			Month 2			Month 3			Month 4			Month 5			Month 6			Month 7			Month 8		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.137	0.054	***	0.076	0.067		0.095	0.068		0.139	0.060	**	0.133	0.062	**	0.114	0.080		0.174	0.068	***	0.099	0.077	
	MSCI	0.194	0.071	***	0.166	0.074	**	0.180	0.074	**	0.182	0.075	**	0.174	0.081	**	0.202	0.084	**	0.211	0.086	**	0.201	0.082	**
	Refinitiv	0.183	0.071	***	0.222	0.074	***	0.191	0.074	***	0.212	0.075	***	0.217	0.081	***	0.290	0.084	***	0.249	0.086	***	0.276	0.081	***
	RepRisk	0.333	0.072	***	0.430	0.074	***	0.456	0.074	***	0.419	0.075	***	0.342	0.063	***	0.409	0.066	***	0.360	0.069	***	0.516	0.068	***
	SPGlobal	0.055	0.055		0.033	0.054		0.079	0.076		0.105	0.076		0.072	0.063		0.088	0.067		0.149	0.087	*	0.057	0.076	
	Sustainalytics	0.237	0.071	***	0.290	0.074	***	0.309	0.074	***	0.225	0.076	***	0.217	0.061	***	0.242	0.064	***	0.250	0.067	***	0.234	0.067	***
	TVL	0.634	0.071	***	0.637	0.074	***	0.643	0.074	***	0.560	0.075	***	0.671	0.081	***	0.748	0.084	***	0.515	0.068	***	0.745	0.082	***
	Moody's	0.220	0.071	***	0.209	0.074	***	0.223	0.074	***	0.238	0.075	***	0.245	0.081	***	0.266	0.084	***	0.283	0.085	***	0.308	0.081	***
GBP	ISS	-0.218	0.101	**	-0.223	0.104	**	-0.248	0.109	**	-0.281	0.111	***	-0.227	0.113	**	-0.185	0.108	*	-0.130	0.112		-0.238	0.116	**
	MSCI	-0.318	0.136	**	-0.242	0.132	*	-0.269	0.127	**	-0.263	0.140	*	-0.242	0.157		0.069	0.147		0.047	0.143		0.021	0.136	
	Refinitiv	-0.309	0.132	**	-0.314	0.128	**	-0.340	0.122	***	-0.379	0.136	***	-0.299	0.153	*	-0.231	0.142		-0.169	0.140		-0.244	0.133	*
	RepRisk	0.194	0.138		0.154	0.133		0.249	0.127	**	0.289	0.140	**	0.388	0.155	**	0.401	0.141	***	0.282	0.140	**	0.395	0.132	***
	SPGlobal	-0.313	0.133	**	-0.280	0.129	**	-0.324	0.123	***	-0.345	0.137	***	-0.320	0.153	**	-0.261	0.141	*	-0.180	0.140		-0.278	0.133	**
	Sustainalytics	-0.457	0.133	***	-0.490	0.128	***	-0.503	0.122	***	-0.585	0.136	***	-0.449	0.153	***	-0.256	0.143	*	-0.225	0.141		-0.302	0.134	**
	TVL	0.705	0.127	***	0.699	0.123	***	0.792	0.117	***	0.706	0.131	***	0.693	0.145	***	0.678	0.132	***	0.484	0.135	***	0.480	0.129	***
	Moody's	-0.252	0.133	*	-0.257	0.129	**	-0.282	0.123	**	-0.329	0.136	**	-0.266	0.153	*	-0.279	0.140	**	-0.214	0.139		-0.298	0.133	**
JPY	ISS	0.143	0.056	***	0.244	0.056	***	0.257	0.052	***	0.230	0.054	***	0.178	0.052	***	0.271	0.055	***	0.321	0.053	***	0.374	0.054	***
	MSCI	0.477	0.068	***	0.510	0.067	***	0.394	0.066	***	0.488	0.069	***	0.519	0.064	***	0.424	0.067	***	0.495	0.068	***	0.533	0.072	***
	Refinitiv	0.104	0.071		0.157	0.068	**	0.170	0.066	***	0.135	0.070	*	0.145	0.064	**	0.184	0.067	***	0.228	0.068	***	0.292	0.070	***
	RepRisk	0.063	0.074		-0.055	0.072		-0.257	0.069	***	-0.257	0.073	***	0.053	0.069		-0.143	0.072	**	0.107	0.074		0.044	0.075	
	SPGlobal	0.330	0.068	***	0.338	0.066	***	0.188	0.066	***	0.316	0.069	***	0.337	0.063	***	0.286	0.067	***	0.344	0.068	***	0.374	0.072	***
	Sustainalytics	0.473	0.069	***	0.407	0.067	***	0.425	0.066	***	0.400	0.070	***	0.373	0.064	***	0.445	0.067	***	0.509	0.068	***	0.585	0.070	***
	TVL	0.707	0.072	***	0.708	0.071	***	0.710	0.070	***	0.741	0.073	***	0.836	0.068	***	0.785	0.071	***	0.870	0.074	***	0.877	0.076	***
	Moody's	0.265	0.068	***	0.290	0.066	***	0.281	0.065	***	0.273	0.069	***	0.293	0.063	***	0.281	0.066	***	0.304	0.068	***	0.355	0.071	***
USD	ISS	0.021	0.041		0.020	0.042		0.000	0.042		0.006	0.043		0.002	0.045		-0.002	0.047		-0.035	0.047		-0.059	0.050	
	MSCI	0.149	0.057	***	0.190	0.060	***	0.124	0.059	**	0.139	0.060	**	0.185	0.057	***	0.150	0.062	**	0.124	0.065	*	0.107	0.067	
	Refinitiv	0.032	0.057		-0.009	0.043		-0.027	0.042		-0.016	0.043		-0.019	0.059		-0.025	0.064		-0.038	0.066		-0.064	0.050	
	RepRisk	0.036	0.057		0.002	0.043		0.146	0.059	**	0.160	0.060	***	0.113	0.058	*	0.064	0.063		0.097	0.066		-0.018	0.068	
	SPGlobal	0.083	0.057		0.144	0.060	**	0.068	0.059		0.083	0.060		0.111	0.057	*	0.117	0.062	*	0.054	0.065		0.067	0.067	
	Sustainalytics	0.168	0.057	***	0.158	0.062	***	0.124	0.059	**	0.105	0.061	*	0.109	0.058	*	0.131	0.048	***	0.042	0.066		0.215	0.068	***
	TVL	0.294	0.057	***	0.387	0.060	***	0.375	0.058	***	0.354	0.059	***	0.414	0.057	***	0.442	0.062	***	0.372	0.065	***	0.456	0.066	***
	Moody's	0.098	0.057	*	0.103	0.060	*	0.077	0.059		0.081	0.059		0.105	0.057	*	0.110	0.062	*	0.088	0.064		0.109	0.066	

Table A6. Lasso IV estimates for month-by-month returns. This table reports estimates of β from the 2SLS regression (29), with $r_{k,t+h}$ now measuring return in month $t+h$. The Lasso procedure is described in Section 3.3. All reported coefficients and standard errors are multiplied by 100. The regressions are run for each rater, whose names are reported in the left column separately for 4 currency regions: the eurozone, the UK, Japan, and the US. Standard errors are clustered by month and GICS sub-industry. *p<0.1; **p<0.05; ***p<0.01.

		IV Lasso																							
		Month 1			Month 2			Month 3			Month 4			Month 5			Month 6			Month 7			Month 8		
Region	Rater	Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std		Coef	Std	
EUR	ISS	0.137	0.054	***	0.104	0.054	*	0.128	0.057	**	0.139	0.060	**	0.133	0.062	**	0.152	0.066	**	0.174	0.068	***	0.142	0.069	**
	MSCI	0.194	0.071	***	0.166	0.074	**	0.180	0.074	**	0.182	0.075	**	0.174	0.081	**	0.202	0.084	**	0.211	0.086	**	0.201	0.082	**
	Refinitiv	0.183	0.071	***	0.168	0.074	**	0.191	0.074	***	0.212	0.075	***	0.217	0.081	***	0.232	0.085	***	0.249	0.086	***	0.224	0.082	***
	RepRisk	0.333	0.072	***	0.430	0.074	***	0.456	0.074	***	0.419	0.075	***	0.405	0.082	***	0.476	0.086	***	0.421	0.088	***	0.579	0.084	***
	SPGlobal	0.103	0.072		0.082	0.076		0.106	0.076		0.131	0.076	*	0.127	0.082		0.144	0.086		0.179	0.087	**	0.155	0.084	*
	Sustainalytics	0.237	0.071	***	0.219	0.074	***	0.237	0.075	***	0.225	0.076	***	0.222	0.081	***	0.247	0.085	***	0.255	0.087	***	0.239	0.083	***
	TVL	0.634	0.071	***	0.637	0.074	***	0.643	0.074	***	0.560	0.075	***	0.623	0.081	***	0.702	0.085	***	0.639	0.086	***	0.704	0.082	***
	Moody's	0.220	0.071	***	0.209	0.074	***	0.223	0.074	***	0.238	0.075	***	0.245	0.081	***	0.266	0.084	***	0.283	0.085	***	0.252	0.082	***
GBP	ISS	-0.239	0.101	**	-0.242	0.104	**	-0.267	0.109	**	-0.294	0.111	***	-0.227	0.113	**	-0.185	0.108	*	-0.130	0.112		-0.238	0.116	**
	MSCI	0.017	0.141		0.077	0.136		0.075	0.131		0.021	0.144		0.021	0.160		0.069	0.147		0.047	0.143		0.021	0.136	
	Refinitiv	-0.328	0.132	**	-0.336	0.127	***	-0.366	0.121	***	-0.404	0.135	***	-0.324	0.152	**	-0.231	0.142		-0.169	0.140		-0.244	0.133	*
	RepRisk	0.194	0.138		0.154	0.133		0.249	0.127	**	0.289	0.140	**	0.388	0.155	**	0.401	0.141	***	0.282	0.140	**	0.395	0.132	***
	SPGlobal	-0.324	0.133	**	-0.292	0.129	**	-0.336	0.123	***	-0.355	0.136	***	-0.328	0.153	**	-0.261	0.141	*	-0.180	0.140		-0.278	0.133	**
	Sustainalytics	-0.340	0.135	***	-0.369	0.129	***	-0.376	0.124	***	-0.477	0.137	***	-0.353	0.155	**	-0.256	0.143	*	-0.225	0.141		-0.302	0.134	**
	TVL	0.705	0.127	***	0.699	0.123	***	0.792	0.117	***	0.706	0.131	***	0.693	0.145	***	0.678	0.132	***	0.484	0.135	***	0.480	0.129	***
	Moody's	-0.330	0.131	***	-0.337	0.127	***	-0.368	0.121	***	-0.401	0.134	***	-0.334	0.151	**	-0.279	0.140	**	-0.214	0.139		-0.298	0.133	**
JPY	ISS	0.187	0.055	***	0.244	0.056	***	0.230	0.053	***	0.230	0.054	***	0.248	0.051	***	0.243	0.055	***	0.292	0.054	***	0.348	0.054	***
	MSCI	0.293	0.068	***	0.386	0.067	***	0.394	0.066	***	0.369	0.069	***	0.393	0.063	***	0.424	0.067	***	0.495	0.068	***	0.533	0.072	***
	Refinitiv	0.128	0.070	*	0.180	0.067	***	0.191	0.065	***	0.158	0.070	**	0.166	0.064	***	0.205	0.067	***	0.250	0.067	***	0.292	0.070	***
	RepRisk	-0.327	0.058	***	-0.373	0.060	***	-0.335	0.063	***	-0.348	0.062	***	-0.335	0.059	***	-0.253	0.057	***	-0.063	0.084		-0.110	0.088	
	SPGlobal	0.173	0.070	**	0.214	0.068	***	0.225	0.065	***	0.202	0.070	***	0.221	0.064	***	0.243	0.066	***	0.292	0.067	***	0.323	0.070	***
	Sustainalytics	0.333	0.070	***	0.407	0.067	***	0.425	0.066	***	0.400	0.070	***	0.439	0.063	***	0.445	0.067	***	0.509	0.068	***	0.585	0.070	***
	TVL	0.730	0.072	***	0.728	0.071	***	0.729	0.070	***	0.760	0.073	***	0.852	0.068	***	0.785	0.071	***	0.887	0.074	***	0.891	0.076	***
	Moody's	0.265	0.068	***	0.290	0.066	***	0.281	0.065	***	0.273	0.069	***	0.293	0.063	***	0.281	0.066	***	0.325	0.067	***	0.355	0.071	***
USD	ISS	0.021	0.041		0.020	0.042		0.000	0.042		0.006	0.043		-0.007	0.045		-0.012	0.047		-0.035	0.047		-0.031	0.050	
	MSCI	0.149	0.057	***	0.146	0.061	**	0.124	0.059	**	0.139	0.060	**	0.144	0.058	**	0.129	0.062	**	0.102	0.065		0.082	0.067	
	Refinitiv	0.032	0.057		0.025	0.061		0.003	0.059		0.016	0.060		-0.003	0.059		-0.008	0.063		-0.023	0.066		-0.025	0.069	
	RepRisk	0.062	0.057		0.079	0.061		0.146	0.059	**	0.160	0.060	***	0.217	0.058	***	0.171	0.063	***	0.207	0.065	***	0.126	0.068	*
	SPGlobal	0.083	0.057		0.098	0.060		0.068	0.059		0.083	0.060		0.070	0.058		0.076	0.062		0.054	0.065		0.067	0.067	
	Sustainalytics	0.168	0.057	***	0.158	0.062	***	0.124	0.059	**	0.105	0.061	*	0.109	0.058	*	0.085	0.063		0.042	0.066		0.073	0.068	
	TVL	0.294	0.057	***	0.387	0.060	***	0.375	0.058	***	0.354	0.059	***	0.414	0.057	***	0.418	0.062	***	0.372	0.065	***	0.456	0.066	***
	Moody's	0.098	0.057	*	0.103	0.060	*	0.077	0.059		0.081	0.059		0.073	0.058		0.074	0.062		0.050	0.065		0.064	0.067	

Table A7. Summary of results. This table presents several summary statistics regarding the noise in the rating agencies scores. In total, there are 512 estimations, given by 8 raters, 8 time horizons, 4 regions, and 2 estimates procedures. The top panel indicates counts in the following categories: the total possible coefficients, the coefficients that were actually estimated, the cases where attenuation is present, the non-attenuation cases (but no sign switching), and the sign switching cases, where the 2SLS coefficient has a different sign than the OLS coefficient. The following panels provide average values of the noise-to-signal ratio κ by region, rater, and return horizon. The last panel presents the overall average of the noise-to-signal ratio for every region, rater, and return horizon. Statistics are shown in separate columns for all coefficients, for those obtained by Pruned IV and for those obtained by Lasso IV.

	All	Pruned IV	Lasso IV
Panel A: Estimation statistics			
Total Possible Coefficients	512	256	256
Estimated Coefficients	427	252	175
Attenuation Cases	302	199	103
Non-Attenuations Cases	92	40	52
Sign switching	33	13	20
Panel B: Noise-to-signal ratio, average by region (%)			
EUR	75.42	74.89	76.98
GBP	58.94	58.15	60.15
JPY	53.61	57.97	45.35
USD	63.20	64.94	59.90
Panel C: Noise-to-signal ratio, average by ESG rater (%)			
ISS	18.22	19.42	16.85
MSCI	57.74	57.72	57.85
Refinitiv	68.43	66.91	71.63
RepRisk	43.04	41.98	49.79
S&P Global	75.96	76.11	75.64
Sustainalytics	54.46	60.26	47.33
TVL	90.73	90.57	91.03
Moody's	65.57	68.79	60.75
Panel D: Noise-to-signal ratio, average by return horizon (%)			
1-month returns	65.77	65.47	66.08
2-month returns	65.71	67.14	63.71
3-month returns	62.42	65.06	57.16
4-month returns	59.97	62.99	52.57
5-month returns	60.01	63.01	52.91
6-month returns	55.71	57.62	50.74
7-month returns	60.44	61.73	57.35
8-month returns	62.17	64.34	56.64
Panel E: Noise-to-signal ratio, overall average (%)			
Overall Average	61.71	63.35	58.55

A.4 Possible Sources of Noise in ESG Scores

An ESG score produced by an ESG rating agency is an aggregate of many indicators measuring a variety of attributes, some of which might be unrelated to each other (such as CO2 emissions and labor practices). The ESG attribute Y_t in our model is therefore an aggregate of a multidimensional variable. In this section, we explain how to think about noise and our noise-correction procedure in this context.

Assume that ESG rating agencies compute the scores as a weighted average of many indicators, corresponding to disaggregated ESG attributes (e.g., CO2 emissions, labor practices):

$$s_{t,i} = \sum_{a \in \{1,n\}} w_{a,i} \cdot I_{a,t,i}, \tag{47}$$

where i indexes ESG rating agencies, a indexes attributes that the agency considers, $I_{a,t,i}$ is a measure of attribute a by rater i,[20] and $w_{a,i}$ are the weights.

The true value of Y_t is given by a similar construct,

$$Y_t = \sum_{a \in \{1,n\}} w_a^\star \cdot I_{a,t}^\star, \tag{48}$$

where $I_{a,t}^\star$ are the true values of the indicators and $w_a^\star$ are the true weights—i.e., the weights that the representative ESG investor assigns to individual indicators, which reflect her preferences or social preferences.

At this stage, some discussion about these constructs might be useful. Suppose there are two attributes that are important to investors: labor practices and CO2 emissions. For a given firm, the true values of labor treatment and CO2 emissions are denoted by $I_{a,t}^\star$. As in our model, a rating agency does not observe these true values; it only observes their proxies. For each agency i, those proxies are the indicators $I_{a,t,i}$. For example, an indicator for labor practices could be constructed based on labor turnover as reported by the firm, or the number of complaints in labor courts. Both indicators are correlated with the true value, but they are not identical to it. The difference is the error term. For the case of CO2 emissions, the indicator could be constructed based on the self-reported emissions (which could be noisy due to the self-reporting nature of this data), or industry estimates (such a procedure is typically used to estimate real estate emissions), or imputed data (e.g., a large fraction of emissions reported by Refinitiv is imputed data). For the weights, investors have preferences between labor treatment and emissions, represented by $(w_a^\star, 1 - w_a^\star)$. The rating agency does not observe these weights and needs to estimate them or use their own. The weights a rating agency uses are not identical to the true weights and the difference is assumed to be a random variable.

[20] The indicators $I_{a,t,i}$ are continuous variables. We normalize them so that they are measured on the same scale.

Under our assumptions, it is possible to decompose the measurement error of each rating agency i as follows:

$$s_{t,i} = Y_t + \underbrace{\sum_{a\in\{1,n\}} w_{a,i} \cdot \underbrace{(I_{a,t,i} - I^\star_{a,t})}_{\eta_{I_{a,t,i}}} + \sum_{a\in\{1,n\}} \underbrace{(w_{a,i} - w^\star_a)}_{\eta_{w_{a,i}}} \cdot I^\star_{a,t}}_{\eta_{Y_{t,i}}} . \tag{49}$$

There are two sources of noise in this decomposition: the measurement error at the level of the indicator,

$$I_{a,t,i} = I^\star_{a,t} + \eta_{I_{a,t,i}}, \tag{50}$$

$$E[\eta_{I_{a,t,i}} | I^\star_{a,t}, w^\star_a] = 0, \tag{51}$$

and the discrepancy in the weights:

$$w_{a,i} = w^\star_a + \eta_{w_{a,i}}, \tag{52}$$

$$E[\eta_{w_{a,i}} | I^\star_{a,t}, w^\star_a] = 0. \tag{53}$$

Equation (51) implies that the measurement error in each indicator is mean independent of the true measure and the true weights. In other words, it implies that the difference in the indicators is truly a measurement error. On the other hand, Equation (53) implies that the deviations of the weights assigned by the rating agencies relative to the weights that describe the true preferences are orthogonal to the true indicators and the true weights themselves. In this case, the intuition is that the differences in the weights are a mean zero random variable.

The above equations parallel our representation in (2). Additionally, we need to assume that the errors are classical, which is satisfied when the measurement error of the indicators, and the deviations in the weights of the rating agency from the true weights are independent of the true ESG attribute Y. Formally,

$$E\left[\sum_{a\in\{1,n\}} w_{a,i} \cdot (I_{a,t,i} - I^\star_{a,t}) + \sum_{a\in\{1,n\}} (w_{a,i} - w^\star_a) \cdot I^\star_{a,t} \middle| Y_t \right] = 0. \tag{54}$$

Condition (54) is satisfied if conditions (51) and (53) hold. We therefore now have two sufficient conditions, (51) and (53), that play the same role as our classical errors-in-variables assumption (7) in the main text.

For one rating agency's score to be a valid instrument for that of another rating agency, one needs to impose two further moment restrictions: (i) the errors in each indicator (Equation (50)) and each weight (Equation (52)) are independent across any two rating agencies (as in the independence assumption (9) in the main text); and (ii) these errors are not correlated with the stock market returns (as in the exclusion restriction (8) in the main text).

The discrepancy between the ESG ratings of two agencies can be decomposed as follows:

$$s_{t,i} - s_{t,j} = \sum_{a \in \{1,n\}} \underbrace{(w_{a,i} - w_{a,j}) \cdot \bar{I}_{a,t}}_{\text{Scope and Weight}} + \sum_{a \in \{1,n\}} \underbrace{\bar{w}_a \cdot (I_{a,t,i} - I_{a,t,j})}_{\text{Measurement}}, \tag{55}$$

where

$$\begin{aligned} \bar{w}_a &= \frac{w_{a,i} + w_{a,j}}{2} \\ \bar{I}_{a,t} &= \frac{I_{a,t,i} + I_{a,t,j}}{2}. \end{aligned}$$

The first term in (55) captures the weight and scope discrepancies highlighted in Berg, Kölbel, and Rigobon (2020). A weight discrepancy occurs when rating agencies assign different weights to the same attribute and the a scope discrepancy occurs when one of the agencies disregards a category, assigning it a weight of zero. The second term in (55) is the discrepancy in measurement of the same indicator.

The easiest way to develop an understanding of what the required moment conditions mean in this setting is to study two special cases: pure measurement and pure weights differences.

Assume that the rating agencies only differ in the measurement of the indicators, i.e., their weights are identical to each other and identical to the true weights. In this case, the scores of rating agencies i and j are given by

$$\begin{aligned} s_{t,i} &= Y_t + \sum_{a \in \{1,n\}} w_a^{\star} \cdot (I_{a,t,i} - I_{a,t}^{\star}), \\ s_{t,j} &= Y_t + \sum_{a \in \{1,n\}} w_a^{\star} \cdot (I_{a,t,j} - I_{a,t}^{\star}). \end{aligned}$$

The two rating agencies' scores are correlated through Y_t and we expect them to strongly predict each other. This is our relevance assumption. Another assumption that we need is the independence assumption, which requires that measurement errors are uncorrelated across the rating agencies (an analog of (9)), i.e.,

$$E\left[(I_{a,t,i} - I^{\star}_{a,t}) \cdot (I_{a,t,j} - I^{\star}_{a,t})\right] = E\left[\eta_{I_{a,t,i}} \cdot \eta_{I_{a,t,j}}\right] = 0, \quad \forall i, j. \tag{56}$$

This assumption is natural if the errors in the indicators are purely mistakes that are specific to individual rating agencies. There are circumstances in which it can be violated. For example, two rating agencies may use similar (possibly imputed) data and similar procedures to compute an indicator. Because their models are based on similar principles, it is reasonable to conjecture that the errors in the procedures of some agencies are correlated with each other. The second possible source of failure of independence is that one rating agency's scores are influenced by the scores of another, which makes their errors correlated. Both of these violations would be detected by the Sargan-Hansen OIR test (see Appendix A.5 for a detailed explanation of why correlated errors lead to a rejection of the OIR test).

The second special case is when the measured indicators are all equal to the true indicators and the discrepancy comes exclusively from weight differences. In this case, the scores of ESG rating agencies i and j take a familiar form:

$$s_{t,i} = Y_t + \sum_{a \in \{1,n\}} (w_{a,i} - w^{\star}_a) \cdot I^{\star}_{a,t},$$
$$s_{t,j} = Y_t + \sum_{a \in \{1,n\}} (w_{a,j} - w^{\star}_a) \cdot I^{\star}_{a,t}.$$

As in the previous case, the scores of the two rating agencies are trivially related to each other through Y_t, satisfying the relevance assumption. The main assumption we need to make here is that weight deviations are independent across rating agencies (an analog of (9)), i.e.,

$$E\left[(w_{a,i} - w^{\star}_a) \cdot (w_{a,j} - w^{\star}_a)\right] = E\left[\eta_{w_{a,i}} \cdot \eta_{w_{a,j}}\right] = 0, \quad \forall i, j. \tag{57}$$

Again, this assumption is testable using the OIR test.

Most rating agencies not only measure individual attributes; they also, by providing a rating, reflect their preferences across those attributes. What matters the most to the rating agencies is reflected in the weights they use.This service from the rating agencies is important. Investors may not have a detailed understanding of all ESG-related issues, nor the resources needed to achieve such understanding. ESG rating agencies strive to understand these issues deeply; and the weights they assign to individual attributes represent the preferences of many investors and individuals they have interacted with. Their goal is to ascertain the weights of a representative ESG-conscious investor, what we call $w^{\star}_a$'s. It is quite likely that the assessment of these weights differs across rating agencies. Being able to instrument for these differences is as important as the ability to instrument for the measurement error at the individual attribute level.

Our instrumental variable approach relies on the identifying assumptions presented in this section (or in Equations (7), (8), and (9)) and there are instances in which their violations pose a threat to our identification. First, we are using a linear representation of ESG scores (Equation (47)). Berg, Kölbel, and Rigobon (2020) show that a linear approximation performs very well in and out of sample. However, if the aggregation rules are non-linear, the non-linearity implies a correlation between the measurement error and the true underlying measure and hence conditions (9) and/or (51) will be violated and the errors will not be classical.

The second potential problem is that two rating agencies have correlated errors. This can occur in a range of scenarios, such as rating agencies using similar data sources and procedures, or rating agencies relying on self-reported data that could have been manipulated by the firms. If the rating agencies impute indicators using similar data and similar models (as argued by Christensen, Serafeim, and Sikochi, 2022), this is likely to result in the violation of the independence assumption (equation (9) or (56)). This violation will be detected by the OIR test. Another possibility is when firms strategic behavior (greenwashing) will produce deviations that are common to the rating agencies. For example, if rating agencies rely on self-reported data, and a firm manipulates its announcements, then the rating agencies share the error. However, as long as one rating agency sees through the manipulation (or uses a methodology that does not rely on companies' disclosure), the OIR test will detect the correlated instruments.

Finally, our instrumental variable approach may fail because the measurements are correlated with the stock-return relevant cash-flow innovations (ϵ_t). Therefore, when a rating agency looks at the realized returns to determine the value of a particular indicator, or the weights of a particular attribute, the instruments fail the exogeneity assumption. This violation will be detected by the OIR test.

A.5 What if Errors are Correlated Across Raters?

In this section we show that if a rater's score is influenced by the scores of another rater, leading to a violation of (9), this is diagnosed by the OIR test. For concreteness, suppose that one rater simply follows another, that is,

$$s_{k,t,1} = Y_{k,t} + \eta_{k,t,1},$$
$$s_{k,t,2} = s_{k,t,1} + u_{k,t},$$

where $u_{k,t}$ is an error term. The second rater's score can be expressed as

$$s_{k,t,2} = Y_{k,t} + \underbrace{\eta_{k,t,1} + u_{k,t}}_{=\eta_{k,t,2}}.$$

In this example, errors are correlated because $E[\eta_{k,t,1} \cdot \eta_{k,t,2}] \neq 0$.

Recall our main structural equation: $r_{k,t+1} = a + \beta Y_{k,t} + M_{k,t} + \epsilon_{k,t}$, where $\epsilon_{k,t}$ is the error term. This equation is equivalent to

$$r_{k,t+1} = a + \beta(s_{k,t,1} - \eta_{k,t,1}) + \nu_{k,t} = a + \beta s_{k,t,1} + \underbrace{\nu_{k,t} - \beta\left(\eta_{k,t,2} - u_{k,t}\right)}_{\text{error}}$$

One can see immediately that s_2 is not a valid instrument for s_1 because it is correlated with the error term. The OIR test, which checks specifically for the correlation of an instrument with the error term, is going to diagnose this.

A similar argument applies to using s_1 as an instrument for s_2. The Sargan-Hansen test would fail in this case as well.

A.6 More Granular Sustainability Ratings as Instruments

In this appendix, we perform the same simulation as in Section 6.2, but with one change. We reduce the total number of attributes covered by the raters from $n = 24$ to $n = 8$. This corresponds to estimating our model at an E, S, or G, rather than at an ESG rating level. In this economy, the true rating is $Y = 1/8 \sum_{a=1}^{8} I_a$. Table A8 below parallels Table 8 in the main text.

Table A8. Simulation $3'$: Fewer instruments than sources of noise, $n = 8$. This table presents the results of the same simulation as in Table 8 but with a reduced number of attributes per rating, at most 8. The true value of the coefficient is 0.5. In the first column, we report the coefficient from our baseline OLS regression (32). In the remaining columns, we present 2SLS estimates, in which we vary the number of instruments used in the first stage from 1 to 5 (labeled as IV1,. . .,IV5). In rows, we vary the number of attributes covered by each rating. The rating s_1 that we are instrumenting covers all 8 attributes.

Number of Attributes per Rating		OLS	IV1	IV2	IV3	IV4	IV5
1	Coefficient	0.07	0.681	0.839	0.826	0.686	0.702
	Std Error	0.031	0.28	0.21	0.186	0.171	0.169
	1st stage F-stat		188	169	144	129	105
2	Coefficient	0.07	0.797	0.804	0.786	0.757	0.687
	Std Error	0.031	0.219	0.194	0.175	0.168	0.162
	1st stage F-stat		310	198	164	133	114
3	Coefficient	0.07	0.466	0.503	0.427	0.354	0.378
	Std Error	0.031	0.209	0.179	0.168	0.157	0.145
	1st stage F-stat		341	234	179	154	146
4	Coefficient	0.07	0.814	0.641	0.639	0.621	0.63
	Std Error	0.031	0.209	0.169	0.157	0.152	0.147
	1st stage F-stat		342	264	205	165	140
5	Coefficient	0.07	0.729	0.727	0.707	0.707	0.727
	Std Error	0.031	0.179	0.166	0.15	0.144	0.139
	1st stage F-stat		466	274	226	184	158
6	Coefficient	0.07	0.405	0.452	0.437	0.492	0.46
	Std Error	0.031	0.167	0.148	0.142	0.137	0.135
	1st stage F-stat		540	347	254	206	170
7	Coefficient	0.07	0.636	0.487	0.494	0.498	0.492
	Std Error	0.031	0.173	0.151	0.144	0.137	0.134
	1st stage F-stat		500	333	246	205	172
8	Coefficient	0.07	0.645	0.448	0.477	0.451	0.488
	Std Error	0.031	0.163	0.148	0.139	0.135	0.133
	1st stage F-stat		569	348	265	210	175