ESG Confusion and Stock Returns: Tackling the Problem of Noise

Florian Berg, Julian Kölbel, Anna Pavlova, Roberto Rigobon

October 2023

ESG Investing

- Investors want to support "green" and punish "brown" firms
- How do portfolio managers identify green stocks?
  - They use ESG scores provided by ESG rating agencies
- But ESG ratings are noisy, so?
  - How to estimate the effect of ESG on stock returns when the regressor is noisy?
  - How to use the different ESG scores from the different rating agencies?

ESG Performance and Stock Returns

- We would like to estimate

$$\underbrace{\Delta r_{t+1}}_{\text{Stock Returns}} = a + b \cdot \underbrace{Y_{i,t}}_{\text{True ESG Performance}} + \nu_t$$

- One way is to estimate using the ESG scores from rating agencies

$$\underbrace{\Delta r_{t+1}}_{\text{Stock Returns}} = a + b \cdot \underbrace{s_{i,t}}_{\text{Rating Score}} + \nu_t$$

- But $s_{i,t}$ is measured badly
- This implies that the estimated coefficient (b) is **Biased**
    - How can we improve the estimates?

What do we do?

- Underlying assumption of the approach
  - Even though the ESG scores are not perfectly correlated, they are indeed correlated.
  - The ESG rating agencies claim they are measuring the same source of risk
- We use an IV approach.
  - We use the scores of several rating agencies to instrument for another rating agency.
- Why not use standard approaches?
  - Why not use simple average?
  - Why not use Principal Components?

We Find

- On average, standard regression estimates increase on average by a factor of 2.6
- Some agencies' ESG scores have very large noise-to-signal ratios
- There are big differences between raw rankings of firms and noise-corrected rankings
  - 58.4% of firms move up/down one or more quintile after noise-correction
- While our model predicts a "greenium," we are finding the opposite in our sample
  - Green stock did well in our sample period
  - Consistent with van der Beck (2022), Pastor, Stambaugh and Taylor (2022), Karolyi et al. (2023)

Data: ESG Ratings

- We use data from 8 different ESG rating agencies: ISS, Moody's, MSCI, Refinitiv, RepRisk, S&P Global, Sustainalytics, and Truvalue Labs.
- ESG raters rely on different data sources that qualitatively imply different levels of noise

Source	**CSR Reports**	**Regulatory Filings**	**Media**	**Questionnaires**	**Modelled Data**
Availability	Public Self-reported	Public Self-reported	Public Third-party	Private Self-reported	Private Third-party
Reporting	(Voluntary)	(Mandatory)	(Involuntary)	(Voluntary)	(Involuntary)
Noise Level	Medium	Low	High	Medium	High

Empirical Challenges

Problems that we need to address:

- Noise of each ESG rating agency is different in magnitude and nature
- Noise in the scores might be non-classical
- ESG rating score noises might be related to each other (failing the exclusion restriction)
- Returns and ESG performance might be driven by an omitted variable. Hence, ESG scores could also be driven by the same omitted variable

Our approach:

- Use instrumental variables as opposed to averages, or time series filters
- Rely on OIR tests to check for non-classical and invalid instruments.

Empirical Specification

- Structural model, to be estimated at individual stock level

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

where $r_{k,t+1}$ represent stock returns.
$Y_{k,t}$ is the "true" ESG performance.
$M_{k,t}$ is an omitted variable.

- Regression equivalent

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

where

$$s_{k,t,i} = Y_{k,t} + \eta_{k,t,i}$$

is an imperfect measure of the ESG attribute Y_t and $\nu_{k,t} = M_{k,t} + \epsilon_{k,t} - \eta_{k,t,i} \cdot \beta$

Coefficients and Potential Biases

- The coefficient of interest is β. The OLS estimate is

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

- Identifying assumptions: $s_{k,t,j}$ is a valid instrument for $s_{k,t,i}$ if:
  - Relevance: $Cov(s_{k,t,i}, s_{k,t,j}) \neq 0$
  - Classical Errors: $E[\eta_{k,t,i} | Y_{k,t}] = 0, \quad \forall i.$
  - Independence: $E[\eta_{k,t,i} \cdot \epsilon_{k,t}] = 0$ and $E[\eta_{k,t,i} \cdot M_{k,t}] = 0, \quad \forall i.$
  - Exclusion restriction: $E[\eta_{k,t,i} \cdot \eta_{k,t,j}] = 0 \quad \forall j \neq i$
- The IV estimate is

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

Notice that both estimates are biased due to the omitted variable, but the IV estimate does not suffer from the attenuation bias

1 Month stock returns

Indicator	Rater	OLS			Pruned IV											
		β_{OLS}	StdErr		β_{2SLS}	StdErr		Diff	Coherent IV							Ftest
									IS	MS	Re	SP	Su	TV	Mo	
ESG	ISS	0.009	0.005	*	0.024	0.008	***	+		✓	✓	✓	✓	✓	✗	10253
	MSCI	0.011	0.004	***	0.024	0.008	***	+	✓		✓	✓	✓	✓	✗	4715
	Refinitiv	0.007	0.004	*	0.020	0.007	***	+	✓	✓		✓	✓	✓	✗	8448
	SPGlobal	0.009	0.004	**	0.019	0.007	***	+	✓	✓	✓		✓	✓	✗	8696
	Sustainalytics	0.024	0.006	***	0.027	0.012	**	+	✓	✓	✓	✓		✓	✓	3707
	TVL	0.001	0.004		0.062	0.019	***	+	✓	✓	✓	✓	✓		✗	489
	Moody's	-0.001	0.005		0.020	0.007	***	[+]	✓	✓	✓	✓	✓	✓		11671
									IS	MS	Re	SP			Mo	
E	ISS-E	0.010	0.005	**	0.014	0.007	*	+		✓	✓	✓			✓	13117
	MSCI-E	0.010	0.005	**	0.023	0.013	*	+	✓		✓	✓			✓	3216
	Refinitiv-E	0.006	0.004		0.014	0.007	**	+	✓	✓		✓			✓	11760
	SPGlobal-E	0.012	0.004	***	0.010	0.006		−	✓	✓	✓				✓	14804
	Moody's-E	0.002	0.005		0.017	0.007	***	+	✓	✓	✓	✓				16729
										MS	Re	SP			Mo	
S	MSCI-S	0.005	0.004		0.007	0.016		+			✓	✓			✓	1685
	Refinitiv-S	0.006	0.004		0.004	0.008		−		✓		✓			✓	11865
	SPGlobal-S	0.007	0.004	*	0.003	0.008		−		✓	✓				✓	11065
	Moody's-S	-0.003	0.005		0.016	0.008	*	[+]		✓	✓	✓				13750
									IS	MS	Re	SP			Mo	
G	ISS-G	0.001	0.005		0.011	0.012		+		✓	✓	✓			✓	3117
	MSCI-G	0.008	0.004	**	0.009	0.016		+	✓		✓	✓			✓	1369
	Refinitiv-G	0.005	0.004		0.008	0.010		+	✓	✓		✓			✓	2386
	SPGlobal-G	0.007	0.004	*	-0.001	0.009		[−]	✓	✓	✓				✓	4392
	Moody's-G	-0.005	0.005		0.021	0.011	*	[+]	✓	✓	✓	✓				5575

2 Month stock returns

Indicator	Rater	β_{OLS}	OLS StdErr		β_{2SLS}	Pruned IV StdErr		Diff	Coherent IV IS	MS	Re	SP	Su	TV	Mo	Ftest
ESG	ISS	0.006	0.005		0.021	0.008	***	+		✓	✓	✓	✓	✓	✗	10076
	MSCI	0.012	0.004	***	0.020	0.008	***	+	✓		✓	✓	✓	✓	✗	4637
	Refinitiv	0.005	0.004		0.015	0.007	**	+	✓	✓		✓	✗	✓	✗	10197
	SPGlobal	0.007	0.004	*	0.012	0.007	*	+	✓	✓	✓		✗	✓	✗	10142
	Sustainalytics	0.023	0.006	***	0.023	0.012	*	–	✓	✓	✓	✓		✓	✓	3652
	TVL	0.002	0.004		0.064	0.020	***	+	✓	✓	✓	✓	✓		✗	481
	Moody's	-0.003	0.005		0.017	0.007	**	[+]	✓	✓	✓	✓	✓	✓		11493
									IS	MS	Re	SP	Mo			
E	ISS-E	0.008	0.005		0.011	0.007		+		✓	✓	✓	✓			12890
	MSCI-E	0.009	0.005	*	0.018	0.013		+	✓		✓	✓	✓			3163
	Refinitiv-E	0.004	0.004		0.011	0.007		+	✓	✓		✓	✓			11543
	SPGlobal-E	0.010	0.004	**	0.008	0.007		–	✓	✓	✓		✓			14542
	Moody's-E	0.001	0.005		0.014	0.007	**	+	✓	✓	✓	✓				16445
									MS	Re	SP	Mo				
S	MSCI-S	0.005	0.004		-0.002	0.016		[–]		✓	✓	✓				1666
	Refinitiv-S	0.004	0.004		0.001	0.008		–	✓		✓	✓				11677
	SPGlobal-S	0.005	0.004		-0.001	0.008		[–]	✓	✓		✓				10893
	Moody's-S	-0.005	0.005		0.011	0.008		[+]	✓	✓	✓					13534
									IS	MS	Re	SP	Mo			
G	ISS-G	0.000	0.005		0.007	0.012		[+]		✓	✓	✓	✓			3055
	MSCI-G	0.007	0.004	*	0.006	0.016		–	✓		✓	✓	✓			1341
	Refinitiv-G	0.005	0.004		0.005	0.010		+	✓	✓		✓	✓			2343
	SPGlobal-G	0.005	0.004		-0.002	0.009		[–]	✓	✓	✓		✓			4298
	Moody's-G	-0.005	0.006		0.016	0.011		[+]	✓	✓	✓	✓				5462

Why alternative noise-reduction procedures are less effective?

Simple Average of Rating Agencies' Scores

- If all the errors were perfectly uncorrelated and have the same variance, the simple average is the best index
- The optimal weight should have lower weights on the noisiest variables (that is exactly what the first stage regression does)
- If some of the scores of rating agencies are "invalid", for instance that the errors are correlated with each other or the fundamental, then the simple average includes a misspecified regressor.
- Our IV procedure test for such cases.

Principal Component Analysis

- PCA finds the linear combination of raters scores that maximizes the observed variance
- Therefore, PCA puts the highest weight on the noisiest rating

Multiple Horizons

We evaluate the robustness of the errors-in-variables estimates by changing the dependent variable.

Structural Form

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

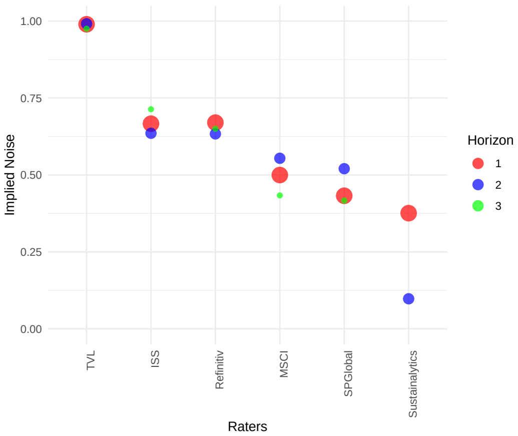
Reduced Form

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

Testable implications:

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

Implied noise in ESG ratings



This figure illustrates the implied noise for stock returns measured over different horizons (from time t to $t+1$ in red, from time $t+1$ to $t+2$ blue, and from time $t+2$ to $t+3$ in green). The vertical axis plots the noise-to-signal ratio κ_i. Raters are sorted by the median observation along the horizontal axis. Observations above one and below zero are not shown as there is no economic interpretation. This results in an exclusion of one observation for Sustainalytics and all three observations for Moody's ESG ratings.

Conclusion

- ESG ratings are noisy.
- Ratings of different rating agencies can be used as instruments.
- Point estimates more than double.
- Attenuation bias is stable across horizons for the same region/rating agency.
- IV is superior to other noise reduction methods.

Backup Slides

Noise for each Rating Agency

The magnitude of the noise in the ESG score used as a regressor can be estimated as

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

where κ_i is the noise-to-signal ratio in the rater i's scores