Are ESG Ratings Noisy for Stock Returns? Evidence from the Thailand Stock Market[1]

Sittisak Leelahanon[a] and Wasin Siwasarit[b]

[a]Faculty of Economics, Thammasat University

[b]Department of Banking and Finance, Chulalongkorn Business School

Corresponding author: wasin@cbs.chula.ac.th

Abstract

We examine the relationship between ESG ratings and stock returns in the Stock Exchange of Thailand (SET), focusing on the measurement-error problem inherent in ESG data. Ratings from Refinitiv, Bloomberg, and S&P Global exhibit pairwise correlations as low as 0.50, indicating that any single score is a noisy proxy for a firm's true ESG quality. This noise induces attenuation bias in standard fixed-effects (FE) regressions, causing ESG coefficients to be biased toward zero. Drawing on the errors-in-variables framework of Berg et al. (2022), we use ratings from two alternative agencies as instrumental variables and estimate a fixed-effect two-stage least squares (FE2SLS) model. Correcting for measurement error yields ESG coefficients that are more than twice the size of FE estimates and that become statistically significant for all three agencies. The results reveal a robust negative relationship between ESG scores and stock returns, consistent with the green premium theory. We also propose and implement a minimum-variance attenuation bias estimator, finding that S&P Global scores are noisiest while Bloomberg scores carry the strongest signal. Governance (G) subscores contain essentially no true information across all three agencies.

[1]We thank Thailand Capital Market Development Fund (CMDF) for supporting S&P Capital IQ data access and participants at the SET Capital Market Research Forum "ESG and Sustainability in Emerging Markets: Shaping Thailand's Future." We gratefully acknowledge Thailand Capital Market Research Institute (CMRI) for the research grant. All errors are our own.

Keywords: ESG ratings, stock returns, measurement error, attenuation bias, instrumental variables, Thailand

JEL Classification: C26, C58, G24, M14

1. Introduction

The relationship between ESG performance and stock returns has attracted substantial academic and practitioner interest. Evidence is mixed: Friede et al. (2015) document a positive ESG–performance link in North America and emerging markets, while others find weak or negative associations in different regions. Latorre et al. (2020) find no ESG effect for Eurostoxx 50 firms, and Garcia et al. (2017) report that superior ESG firms are generally less profitable in BRICS markets. Cayon et al. (2021) document a positive ESG–return correlation for sin companies but a negative one for top- and bottom-quartile non-sin companies in the subsequent year.

A key, underappreciated reason for this disagreement is *rating divergence*: ESG scores from different agencies disagree substantially because they rely on different data sources, estimation models, and weighting schemes (Berg et al., 2022). When a noisy proxy replaces the true underlying construct in a regression, the coefficient is biased toward zero—a phenomenon known as attenuation bias.

We address this problem in the Thai stock market. Our three contributions are as follows. First, we adapt the Berg et al. (2022) errors-in-variables framework to a panel setting and prove consistency of the fixed-effect two-stage least squares (FE2SLS) estimator. Second, we propose a minimum-variance attenuation bias estimator based on a system of seemingly unrelated regressions (SUR) with standard errors from the delta method. Third, we provide the first systematic evidence on ESG measurement quality in Thailand, documenting that governance sub-scores carry near-zero signal across all three rating agencies.

2. Data

We combine ESG ratings from three providers—Refinitiv, Bloomberg, and S&P Global—for 30 SET 50 firms over 2015–2022, with monthly stock returns from SETSMART. Refinitiv and S&P Global scores are on a 0–100 scale; Bloomberg scores range from 0–10. Table 1

presents summary statistics.

Table 1: Descriptive Statistics

Variable	Mean	S.D.	Min	Max
ESG Scores				
Refinitiv ESG	64.82	13.55	17.85	91.82
Bloomberg ESG	3.65	1.38	0.85	6.62
S&P Global ESG	66.90	20.46	14.00	93.00
Financial Variables				
Returns (%)	0.25	2.40	−5.88	16.78
Beta	0.48	0.39	−0.56	2.44
Dividend Yield	3.05	1.94	0.00	10.92
Market-to-Book	3.31	3.02	0.40	15.12
ROA	6.46	4.89	−7.44	26.66
Momentum	0.24	2.53	−6.62	19.41
Volatility	0.08	0.05	0.02	0.29
Asset Growth	10.23	16.34	−14.06	121.01

Note: Returns are average monthly returns (%) from month +1 to +12. Beta estimated from monthly returns, months −60 to −1. Dividend Yield is prior-12-month dividends divided by prior month-end price. Market-to-Book is log(market equity) minus log(book equity). Asset Growth is log growth in total assets. ROA is income before extraordinary items divided by average total assets. Momentum is the return from month −12 to −2. Volatility is the monthly standard deviation from months −12 to −1.

The extent of rating divergence is illustrated in Figure 1. Panel (a) shows that the Refinitiv–Bloomberg pairwise correlation is only 0.498, while panels (b)–(d) reveal individual firms that receive high scores from one agency but low scores from another—a pattern inconsistent with all agencies measuring the same underlying construct.

3. Methodology

Baseline panel model. We estimate a firm fixed-effects regression of stock returns on lagged ESG ratings and a standard set of controls: market beta, dividend yield, market-to-book ratio, ROA, momentum, volatility, and asset growth.

Errors-in-variables correction. Because every observed ESG score equals the true unobserved ESG quality plus an idiosyncratic measurement error, ordinary fixed-effects

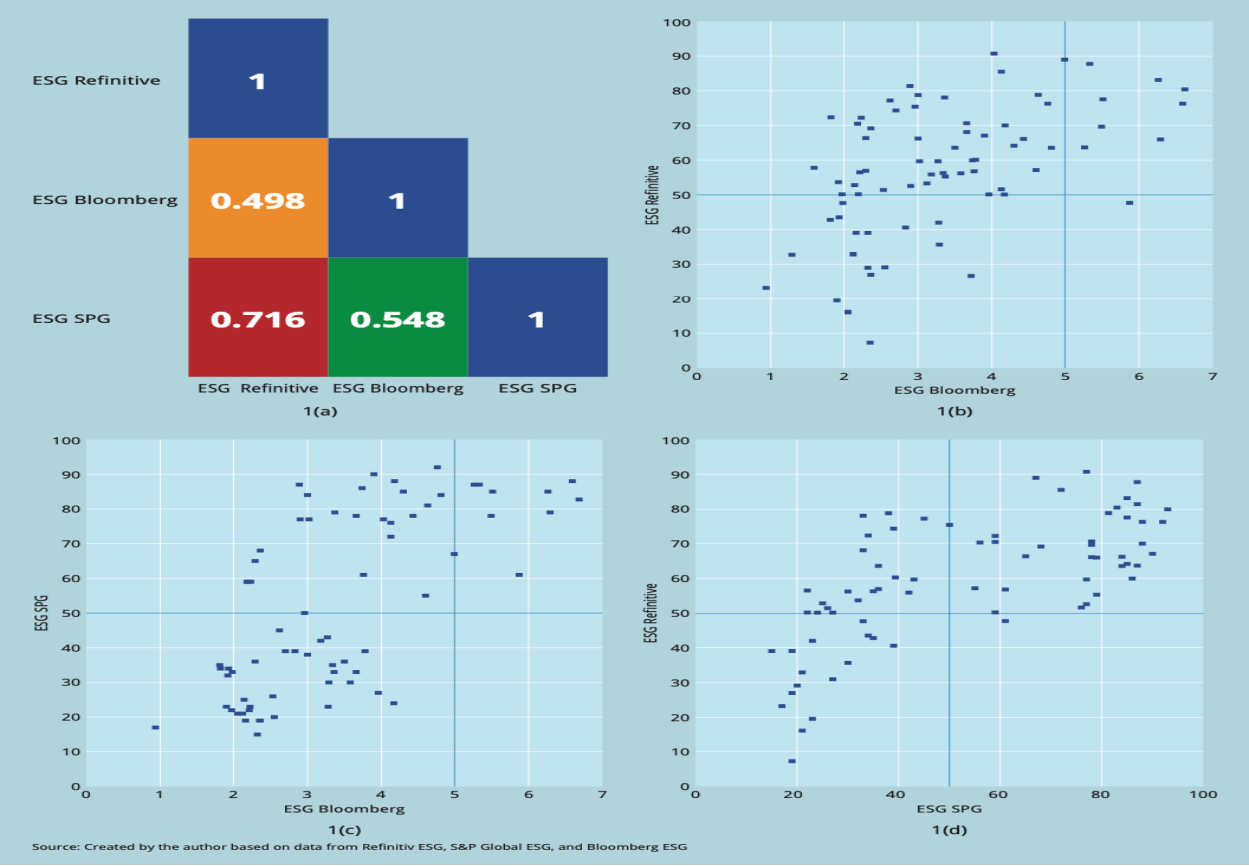

Figure 1: Correlation matrix of ESG scores (panel a) and pairwise scatter plots across agencies in 2022 (panels b–d).

estimation is inconsistent—the coefficient is attenuated toward zero by the factor $\theta = 1 - \sigma_\varepsilon^2/\sigma_x^2 \in (0, 1)$, where σ_ε^2 is the noise variance and σ_x^2 is the total variance of the observed score. We correct for this bias using FE2SLS: ratings from the other two agencies serve as instruments because (i) they correlate with true ESG quality (relevance) and (ii) their idiosyncratic noise components are uncorrelated with those of the instrumented score (exogeneity, given independent measurement processes). Instrument validity is tested via the overidentifying restrictions (Sargan–Hansen) test (Sargan, 1958; Hansen, 1982).

Attenuation bias estimation. To quantify noise levels, we regress each instrument on the instrumented agency's score in a fixed-effects SUR system and show that the slope coefficients converge in probability to θ. A minimum-variance weighted-average esti-

mator is formed from the two slope estimates, with optimal weights derived from the SUR covariance matrix and standard errors obtained by the delta method.

4. Empirical Results

Main results. Table 2 compares FE and FE2SLS estimates. The FE coefficients are negative but weak: Refinitiv is only marginally significant, Bloomberg is significant at 1%, and S&P Global is insignificant. After correcting for measurement error, all three coefficients roughly double in magnitude and gain statistical significance—Refinitiv at 1%, S&P Global at 5%, and Bloomberg at 10%. Overidentifying restrictions tests confirm instrument validity at the 10% level for all three specifications. The negative sign is consistent with the green premium theory: stocks with better ESG profiles attract greater investor demand, trade at higher prices, and consequently earn lower expected returns (Heinkel et al., 2001; Pástor et al., 2021).

Table 2: Stock Returns and ESG Ratings: FE vs. FE2SLS

	Refinitiv		Bloomberg		S&P Global	
	FE	FE2SLS	FE	FE2SLS	FE	FE2SLS
ESG Score	-0.368^{*}	-0.976^{***}	-0.823^{***}	-0.735^{*}	-0.236	-0.762^{**}
	(0.213)	(0.356)	(0.242)	(0.410)	(0.159)	(0.300)
Observations	203	200	209	200	205	200
R^2	0.256	0.209	0.299	0.285	0.250	0.203
OIR Statistic		1.78		0.03		3.06
p-value		0.182		0.856		0.080

Note: Control variables (beta, dividend yield, market-to-book, ROA, momentum, volatility, asset growth) and firm fixed effects are included but not reported. Standard errors in parentheses. ***, **, * denote significance at 1%, 5%, 10%. FE2SLS instruments the agency's score with the scores of the other two agencies. OIR is the Sargan–Hansen overidentification test statistic.

Attenuation bias. The estimated attenuation factors ($\hat{\theta}$) reveal that S&P Global scores are noisiest ($\hat{\theta} \approx 0.33$, implying about two-thirds noise), while Bloomberg scores are cleanest ($\hat{\theta} \approx 0.64$). Refinitiv falls in between ($\hat{\theta} \approx 0.52$). These rankings are stable

Table 3: Attenuation Bias of Each ESG Rating Score

ESG Scores	Refinitiv	Bloomberg	S&P Global
Control variables	0.5251***	0.6384***	0.3454***
are excluded	(0.0459)	(0.0603)	(0.035)
Control variables	0.5091***	0.6439***	0.3241***
are included	(0.043)	(0.0607)	(0.0344)

Note: This table reports attenuation biases of each ESG rating score. The numbers in parentheses are standard errors. ***, **, * denote statistically significance at level 0.01, 0.05, 0.1, respectively.

whether or not control variables are included, supporting robustness. Decomposing by E, S, and G sub-scores, we find that governance (G) attenuation factors are near zero and statistically insignificant across all agencies, indicating that G sub-scores contain virtually no true governance signal.

Robustness. Using separate E, S, and G sub-scores from the other agencies as instruments (rather than aggregate ESG scores) yields quantitatively similar coefficients and maintains instrument validity (all OIR p-values > 0.18). Replacing aggregate ESG with individual E, S, G scores in the return equation produces consistent results for E and S, while G scores remain uninformative.

5. Conclusion

ESG data in Thailand are substantially noisy. Standard fixed-effects regressions mask a genuine negative relationship between ESG quality and stock returns because measurement error attenuates the estimated coefficients toward zero. Correcting for this bias via FE2SLS—using alternative agencies' scores as instruments—recovers coefficients that are more than twice as large and statistically significant. Practitioners should not rely on a single ESG rating: cross-referencing multiple providers and applying instrumental variable methods produces stronger empirical results and more reliable inferences about ESG's role in asset pricing.

Our empirical results come with several limitations. First, our regressions rely on a short time series combined with unbalanced cross-sectional data, which serves as the pri-

mary source of variation. This limitation stems from the nature of ESG ratings, as most rating agencies update their scores at most once a year. Moreover, the available ESG data largely correspond to a period in which the Thai stock market was moving in a "sideway-down" pattern following the COVID-19 pandemic. This market condition may influence our findings and potentially affect the structural interpretation of the coefficients. To perform more rigorous robustness checks, a much longer time series would be required than what is currently available.

Second, this study mainly relies on each individual ESG rating, which raises concerns about whether a single rating can adequately capture the true effect of ESG performance. To address this issue, future research should explore methodologies for constructing a new ESG composite score. Such a score could be calculated as a weighted average of indicators from multiple rating agencies and linked to disaggregated ESG attributes (e.g., carbon emissions, labor practices).

Third, in addition to examining the relationship between ESG performance and stock returns, future work should also investigate how ESG scores relate to stock volatility. A growing body of literature suggests that firms with stronger ESG performance may face lower risk, improved stability, or reduced downside volatility. Extending the analysis in this direction would provide a more comprehensive understanding of how ESG factors shape not only expected returns but also the variability of asset prices.

Lastly, in the context of Thailand, there remains a limitation in studying the asset-pricing and portfolio implications of a key obstacle to sustainable investing: uncertainty surrounding corporate ESG profiles. Theory suggests that under ESG uncertainty, market risk premia increase, investor demand for stocks declines, and both CAPM alpha and effective beta rise—while the negative ESG-alpha relationship becomes weaker. These unanswered questions offer rich opportunities for future research in the Thai market.

References

Berg, F., Kölbel, J. F., and Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.

Cayon, E., Thorp, S., and Wu, E. (2021). Frontier and sin markets: ESG performance and equity returns. *Journal of Sustainable Finance & Investment*, 11(2), 154–174.

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

Garcia, A. S., Mendes-Da-Silva, W., and Orsato, R. J. (2017). Sensitive industries produce better ESG performance: Evidence from emerging markets. *Journal of Cleaner Production*, 150, 135–147.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4), 1029–1054.

Heinkel, R., Kraus, A., and Zechner, J. (2001). The effect of green investment on corporate behavior. *Journal of Financial and Quantitative Analysis*, 36(4), 431–449.

Latorre, V., Saiz-Alvarez, J. M., and Cano, J. A. (2020). Does ESG affect the financial performance of Eurostoxx50 firms? *Corporate Social Responsibility and Environmental Management*, 27(6), 2651–2663.

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

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, 26(3), 393–415.