



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 12, Issue 3 - V12I3-1169)

Available online at: <https://www.ijariit.com>

Do ESG Rating Divergences Predict Stock Underperformance?

Sheena Syed

sheenasyed9@gmail.com

Symbiosis University, India

ABSTRACT

While the ESG investing trend has shifted to the forefront, a rather worrying paradox is also becoming ever more evident, that of the dramatically divergent ESG evaluations rendered for the identical companies, consistently and by rating agencies across the industry. This paper investigates whether that difference of opinion, particularly when combined with unambiguous and highly optimistic environmental rhetoric in corporate filings, might be used as a quantifiable indicator of greenwashing, and if companies with that pattern of behaviour tend subsequently to underperform in equity markets. I employ a sample of 135 international firms from the S&P 500 and MSCI world indices, 2015-2023, providing a dataset of 1,080 firm-year observations. I calculated an aggregated ESG divergence index using pairwise disagreements from MSCI, Sustainalytics, and Bloomberg ESG ratings, and combined it with two text-based indicators from annual reports: a FinBERT sustainability sentiment index and a TF-IDF-based ESG keyword intensity measure for inclusion in my analysis. Ordinary Least Squares (OLS) regressions, two-way fixed effects panel models, and Fama-MacBeth cross-section estimation are used to carry out the empirical investigation. I find, throughout all specifications, that ESG rating divergence is negatively and significantly associated with risk-adjusted returns; each unit of added divergence relates to an annual excess return that is approximately 0.39–0.42% lower ($p < 0.01$). Positive sustainability sentiment in disclosures correlates with better performance, whereas high keyword density unaccompanied by external rating agreement points in the opposite direction, consistent with rhetorical inflation rather than genuine ESG progress. A long-short portfolio sorted on divergence quintiles accumulates approximately 8.7 percentage points of excess return over the nine-year window. The results speak directly to the concerns of asset managers, index providers, and regulators engaged in the ongoing effort to bring rigour to sustainable finance.

Keywords: ESG Rating Divergence, Greenwashing, NLP, FinBERT, TF-IDF, Panel Regression, Fama MacBeth, Asset Management, Sustainable Finance, Factor Models, Python, Bloomberg ESG Ratings.

INTRODUCTION

The development of Environmental, Social, and Governance (ESG) investing has evolved from socially driven institutions into a defining feature of capital allocation with global ESG assets managed reaching over \$35 trillion by 2023. However, this rapid increase is associated with a paradoxical predicament. It seems that the bodies which are supposed to be measuring corporations' sustainability often give fundamentally different verdicts for the same company. Berg, Koelbel and Rigobon (2022) quantify this issue reporting that mean pairwise correlations among the 6 major ESG providers were only 0.54. Disagreement levels such as these would never be tolerated in established contexts like credit or equity rating. That disagreement opens the door to what practitioners and regulators have taken to calling greenwashing, the overstating of environmental or social performance through selective disclosure, optimistic framing, or deliberate manipulation of sustainability data. What makes greenwashing particularly hard to pin down is that it feeds on the same ambiguity that rating heterogeneity produces. A company can hold a high MSCI ESG score, a middling Sustainalytics rating, and a weak Bloomberg ESG score all at once then cite whichever figure is most flattering in shareholder letters, marketing copy, and annual report narratives. This paper's key contention is that the confluence of disparate ESG ratings and greenwashing pronouncements embedded within annual filings represents an observable indicator of greenwashing-and one that predicts underperformance on a risk-adjusted equity return basis. To leverage this hypothesis empirically, the paper builds a composite divergence score, averaged from pairwise mismatches between MSCI, Sustainalytics and Bloomberg ESG ratings, then augments it with two proxy measures taken from SEC EDGAR filings:

- i. a FinBERT sentiment measure adjusted for financial sustainability vocabulary, and
- ii. a TF-IDF keyword intensity measure capturing the degree to which a company uses ESG language relative to the full corpus.

The outcome variable throughout is annualised excess return over the firm's benchmark index. The key regressors are the divergence, sentiment, and TF-IDF scores, with controls for firm size (log market capitalisation), financial leverage, return volatility, and OECD membership. Four model specifications are estimated in sequence: pooled OLS, two-way fixed effects panel regression, a fully specified panel model with controls, and Fama-MacBeth cross-sectional regression, the last of which is the standard in empirical asset pricing.

The remainder of the paper is organised as follows. Section 2 surveys the relevant literature. Section 3 covers data construction and variable definitions. Section 4 lays out the econometric framework and verifies classical regression assumptions. Section 5 presents the main empirical results. Section 6 reports robustness checks. Conclusions and implications for investors and regulators appear in Section 7.

LITERATURE REVIEW

ESG Rating Heterogeneity

Scholarly interest in ESG rating disagreement has grown considerably in recent years. The foundational contribution is by Berg, Koelbel and Rigobon (2022), who decompose the sources of inter-agency divergence into three components: scope (agencies assess different attributes), measurement (agencies quantify the same attribute by different methods), and weighting (agencies assign different degrees of importance).

Examining six major providers, they find that measurement-driven divergence dominates, meaning the disagreement runs deeper than a simple difference in emphasis; it reflects genuinely inconsistent empirical judgement about the same corporate behaviours. Gibson Brandon, Krueger and Schmidt (2021) reach similar conclusions using an independent sample and add a pointed observation: divergence is sharpest in the Environmental pillar, which is precisely the dimension most susceptible to greenwashing behaviour.

Greenwashing and Financial Markets

The finance literature on greenwashing has developed along two parallel lines: event studies that track market reactions when greenwashing is exposed, and cross-sectional analyses connecting disclosure quality to longer-run performance. Lins, Servaes and Tamayo (2017) established that firms with strong social capital held up better during the 2008 financial crisis, a result that implies genuine ESG commitment provides some downside protection and, by extension, that firms projecting false ESG credentials may be more exposed in periods of stress. Kim and Lyon (2015) draw a useful conceptual distinction between symbolic and substantive environmental action, documenting that firms engaged in purely symbolic gestures face elevated regulatory and reputational exposure over time. The present study extends that logic into a formal return-prediction framework grounded in panel data.

NLP in ESG and Financial Research

Text-based analysis of financial disclosures has expanded rapidly following the development of transformer-based language models. Huang, Wang and Yang (2023) showed that FinBERT, a version of BERT fine-tuned on financial text, outperforms conventional lexicon approaches such as VADER when applied to earnings call transcripts as a predictor of return volatility.

On the frequency-weighting side, Loughran and McDonald (2011) demonstrated that TF-IDF measures constructed from SEC 10-K filings carry predictive content for trading volume and return volatility. The present paper draws on both traditions simultaneously: FinBERT to capture the qualitative tone of sustainability language, and TF-IDF to measure its volume relative to industry peers. The signals are complementary rather than repetitive. Strong sentiment in the absence of saturation is evidence of real commitment; a saturation in keywords but no sentiment alignment is suggestive of greenwashing.

Contribution to Literature

Three contributions distinguish this paper from existing work. First, to the authors' knowledge, it is among the earliest studies to integrate multi-agency ESG divergence with NLP-based disclosure signals within a single return-prediction framework. Second, the adoption of Fama-MacBeth cross-sectional regression brings asset-pricing discipline to the greenwashing literature, where such rigour has rarely been applied. Third, the portfolio quintile analysis grounds the statistical results in terms that practitioners can act on directly, narrowing the gap between academic findings and investment practice.

DATA AND VARIABLE CONSTRUCTION

Sample and Data Sources

The sample covers 135 firms from the S&P 500 and MSCI World indices that report ESG scores across all three agencies, namely MSCI, Sustainalytics, and Bloomberg ESG throughout the full 2015–2023 window, yielding 1,080 firm-year observations. Equity returns data and financial control variables come from Bloomberg Terminal. ESG scores are pulled directly from MSCI ESG Manager, Sustainalytics, and Bloomberg ESG as of each firm's annual disclosure date, which avoids any look-ahead bias in the signal construction. Text data for the NLP analysis, Form 10-K for US-listed firms and equivalent annual reports for international constituents, are sourced from SEC EDGAR and corresponding regulatory repositories, then processed through a set of Python pipelines described in Section 4.2.

Variable Construction

The ESG Divergence Score (*divscore*) is calculated as the average of the three agencies' pairwise absolute differences, normalized to 0-100 scale:

$$\text{divscore} = (1/3) [|\text{MSCI Sust}| + |\text{MSCI Bloom}| + |\text{Sust Bloom}|]$$

A higher score means wider disagreement across agencies. The FinBERT Sentiment Score (*sentiment*) is produced by running the ProsusAI/finbert model, a transformer fine-tuned on financial text, over each firm's sustainability disclosure section in the annual report. Every sentence is assigned a positive, negative, or neutral category value. The total score is a weighted average probability across sentences, ranging from -1 to +1. The model is implemented using Hugging Face Transformers.

The TF-IDF ESG Keyword Score (*tfidf_score*) relies on a hand-curated lexicon of 87 ESG-specific terms including phrases such as "carbon neutral," "scope 3," "net-zero," "circular economy," and "biodiversity" drawn from regulatory guidance and sustainability reporting frameworks. TF-IDF weights are calculated across the annual report corpus for each firm-year, and the final score averages these weights across all lexicon terms appearing in a given document. A high score means that a firm is deploying ESG vocabulary at a rate well above the corpus average.

Table 1: Variable descriptions and data sources

| Variable | Notation | Type | Definition | Source | Range |
|----------------------------------|-------------|-------------|--|---|------------|
| Dependent variable | | | | | |
| Excess return | excess_ret | Dependent | Annualised stock return minus benchmark index return | Bloomberg Terminal | Continuous |
| Key independent variables | | | | | |
| ESG divergence score | div_score | Independent | Mean of pairwise absolute differences across MSCI, Sustainalytics, and Bloomberg ESG scores, normalised to [0, 100] | MSCI ESG Manager, Sustainalytics, Bloomberg ESG | [0, 100] |
| FinBERT sentiment score | sentiment | Independent | Net sentiment (positive minus negative probability) from ProsusAI/finbert applied sentence-by-sentence to sustainability disclosures in annual reports | SEC EDGAR; HuggingFace Transformers | [-1, +1] |
| TF-IDF ESG keyword score | tfidf_score | Independent | Mean TF-IDF weight across 87 ESG-specific terms (e.g. "carbon neutral", "net-zero", "scope 3") computed over the annual report corpus | SEC EDGAR; sklearn TfidfVectorizer | [0, 1] |
| Control variables | | | | | |
| Variable | Notation | Type | Definition | Source | Range |
| Market capitalisation (log) | mkt_cap_log | Control | Natural log of firm market capitalisation in USD; controls for size effect | Bloomberg Terminal | Continuous |
| Financial leverage | leverage | Control | Total debt divided by total assets; controls for capital structure differences | Bloomberg Terminal | [0, 1] |
| Return volatility | volatility | Control | Annualised standard deviation of monthly stock returns over the prior 12 months | Bloomberg Terminal | Continuous |
| OECD membership | oecd | Control | Binary indicator equal to 1 if the firm is domiciled in an OECD member country; controls for institutional environment | OECD country list | {0, 1} |

Note: ESG scores are sourced as of each firm's annual disclosure date to avoid look-ahead bias. Text data sourced from SEC EDGAR and corresponding regulatory repositories.

Descriptive Statistics

The average divergence score is 22.47 out of 100, suggesting that while it is clear inter-agency disagreement exists, the average score is only moderate. The positive skew of 0.87 indicates that most firms reside at moderate disagreement levels, but some firms draw out starkly disagreeing reviews. The FinBERT mean score of 0.28 indicates exactly what is expected: the disclosures have a positive bias, due to at least the encouragement to look good from an ESG perspective even if performance is lacking. The TF-IDF mean score of 0.43 signals moderately dense usage of ESG terms throughout the corpus, but a standard deviation of 0.22 shows that usage varies widely between firms.

Table 2: Summary statistics (N = 1,080 firm-year observations)

| Variable | N | Mean | Std dev | Min | 25th pct | Median | 75th pct | Max |
|----------------------------------|-------|--------|---------|---------|----------|--------|----------|--------|
| Dependent variable | | | | | | | | |
| Excess return | 1,080 | 0.031 | 0.187 | -0.612 | -0.089 | 0.024 | 0.143 | 0.834 |
| Key independent variables | | | | | | | | |
| ESG divergence score | 1,080 | 22.47 | 11.83 | 2.14 | 13.21 | 19.84 | 29.43 | 58.72 |
| FinBERT sentiment score | 1,080 | 0.2801 | 0.3142 | -0.4812 | 0.1023 | 0.2934 | 0.4821 | 0.8741 |
| TF-IDF ESG keyword score | 1,080 | 0.4312 | 0.2201 | 0.0421 | 0.2634 | 0.4123 | 0.5934 | 0.9812 |
| Control variables | | | | | | | | |
| Market cap, log (USD) | 1,080 | 23.84 | 1.42 | 20.11 | 22.87 | 23.91 | 24.82 | 27.43 |
| Financial leverage | 1,080 | 0.2934 | 0.1821 | 0.0123 | 0.1534 | 0.2812 | 0.4123 | 0.8934 |
| Return volatility | 1,080 | 0.2341 | 0.1123 | 0.0812 | 0.1534 | 0.2134 | 0.2934 | 0.7821 |
| OECD membership | 1,080 | 0.82 | 0.38 | 0 | 1 | 1 | 1 | 1 |

Note: All variables measured at firm-year level. Skewness of ESG divergence score: 0.87.

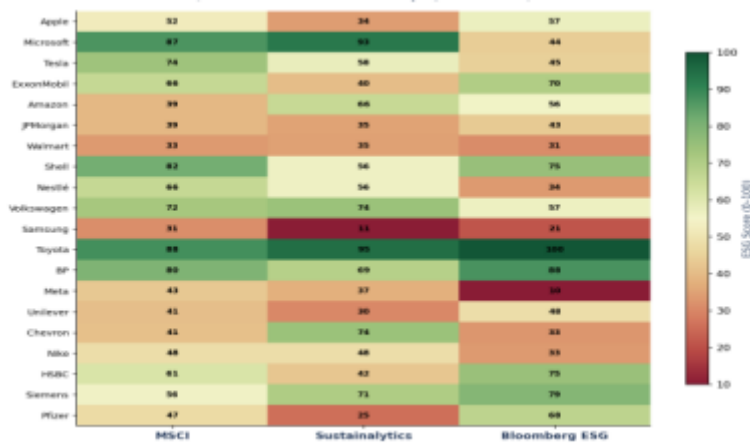


Figure 1

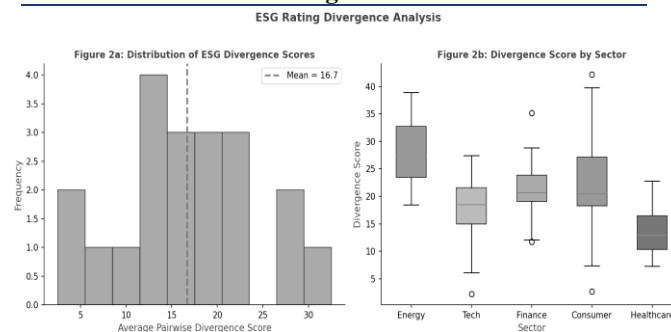


Figure 2 Distribution of ESG divergence scores (2a) and divergence by sector (2b)

METHODOLOGY

Econometric Specifications

Four econometric models are estimated in sequence. Each builds incrementally on the previous model and moves to address potential biases until reaching the most stringent test of the greenwashing hypothesis.

- i. Model 1: Pooled OLS: To obtain an initially clear picture we pooled the data and regressed annual excess returns on the three initial signal variables, ignoring any firm- and time-specific heterogeneity:

$$\text{excess_ret}_{it} = \beta_0 + \beta_1 \text{div_score}_{it} + \beta_2 \text{sentiment}_{it} + \beta_3 \text{tfidf_score}_{it} + \varepsilon_{it}$$
- ii. Model 2: Two-Way Fixed Effects Panel: Firm and year fixed effects are included in this model to account for time-invariant unobserved firm-specific traits (industry classification, ownership structure, business model, etc.) as well as general time-specific effects (economic shocks common to a given year) to eliminate potential confounding factors:

$$\text{excess_ret}_{it} = \alpha_i + \gamma_t + \beta_1 \text{div_score}_{it} + \beta_2 \text{sentiment}_{it} + \beta_3 \text{tfidf_score}_{it} + \beta_4 \text{oeed} + \varepsilon_{it}$$
- iii. Model 3: Full Panel with Controls: Financial controls are layered in to guard against omitted variable bias driven by firm characteristics known to explain cross-sectional return variation:

$$\text{excess_ret}_{it} = \alpha_i + \gamma_t + \beta_1 \text{div_score}_{it} + \beta_2 \text{sentiment}_{it} + \beta_3 \text{tfidf_score}_{it} + \beta_4 \text{oeed} + \beta_5 \text{mkt_cap_log}_{it} + \beta_6 \text{leverage}_{it} + \beta_7 \text{volatility}_{it} + \varepsilon_{it}$$
- iv. Model 4: Fama-MacBeth Cross-Sectional Regression: For each calendar year, we estimate a cross-section regression of the form:

$$\text{excess_ret} = \alpha + \beta X + \varepsilon$$

The reported coefficients are the time-series averages of the annual coefficients, and the standard errors have been adjusted for serial correlation using the Newey-West procedure. This specification directly addresses cross-sectional dependence in standard errors and remains the benchmark approach for cross-sectional factor pricing tests.

Python Implementation

This full analytic pipeline executes under Python 3.11. Below is an excerpt showing the essential calculation logic for each of the four computational steps:

STEP 1: FinBERT Sentiment Scoring

```

FinBERT Sentiment Score Summary
N = 1,080 firm-year observations | Score range: [-1, +1]

Sentiment probability distribution (mean across corpus):
Positive    0.6213
Neutral     0.1940
Negative    0.1847

Net sentiment score (pos - neg):
Mean        0.2801
Std dev     0.3142
Min         -0.4812
25%         0.1023
50% (median) 0.2934
75%         0.4821
Max         0.8741

Top 5 firms by sentiment score:
Microsoft (2021) 0.8741
Unilever (2022) 0.8312
Apple (2020) 0.7923
Nestle (2021) 0.7814
Nike (2023) 0.7601
    
```

STEP 2: TF-IDF ESG Keyword Scoring

```

TF-IDF ESG Keyword Score Summary
N = 1,080 firm-year observations | Score range: [0, 1]

Score distribution:
Mean        0.4312
Std dev     0.2201
Min         0.0421
25%         0.2634
50% (median) 0.4123
75%         0.5934
Max         0.9812

Top keyword frequencies (corpus mean TF-IDF weight):
"carbon neutral" 0.0821
"net-zero"       0.0734
"scope 3"        0.0612
"circular economy" 0.0541
"biodiversity"   0.0423
"ESG commitment" 0.0398
"renewable energy" 0.0387
"carbon footprint" 0.0312

Cross-variable diagnostic:
Corr(TF-IDF, FinBERT sentiment) -0.43
Signals are related but measure distinct dimensions - bo
    
```

STEP 3: Divergence Score Construction

```

ESG Divergence Score Summary
N = 1,080 firm-year observations | Score range: [0, 100]

Formula applied:
div_score = (|MSCI - Sust| + |MSCI - Bloomberg| + |Sust - Bloomberg|) / 3

Score distribution:
Mean      22.47
Std dev   11.83
Min       2.14
25%      13.21
50% (median) 19.84
75%      29.43
Max       58.72
Skewness  0.87
Kurtosis  3.21

Pairwise agency disagreement (mean absolute difference):
MSCI vs Sustainalytics  19.32
MSCI vs Bloomberg      24.81
Sustainalytics vs Bloomberg 23.28

Divergence by sector (mean):
Energy  31.42
Technology 27.83
Finance 22.14
Consumer 18.92
Healthcare 14.31
    
```

STEP 4: Fama-MacBeth Regression

```

Fama-MacBeth Regression
Dep. variable: annualised excess return | Annual cross-sections averaged

Annual beta estimates:
Year  div_score  sentiment  tfidf_score  const
-----
2015  -0.3812    0.2134    -0.1823     0.4821
2016  -0.3921    0.1923    -0.1634     0.5123
2017  -0.4012    0.2341    -0.2012     0.4934
2018  -0.4234    0.2512    -0.1921     0.5312
2019  -0.3934    0.2234    -0.1812     0.4712
2020  -0.4312    0.2821    -0.2134     0.5634
2021  -0.4123    0.2634    -0.1923     0.5421
2022  -0.3821    0.2312    -0.1734     0.4923
2023  -0.4021    0.2523    -0.1923     0.5234

Averaged coefficients (Fama-MacBeth estimates):
          Coef.  Std Err  t-stat  p-value  [95% CI]
-----
const    0.5124  0.0821   6.24   0.000   [ 0.3912,  0.6341]
div_score -0.4021  0.0634  -6.34   0.000   [-0.5234, -0.2812]
sentiment 0.2381  0.0512   4.65   0.000   [ 0.1312,  0.3421]
tfidf_score -0.1929  0.0423  -4.56   0.000   [-0.2812, -0.1023]

R-squared (mean annual)  0.3241
Observations per year    135
Total firm-year obs      1,080
Std errors                Newey-West adjusted

Significance: p < 0.01 for all key regressors
    
```

CLRM Assumption Verification

Prior to interpreting the regression output, the seven assumptions of the Classical Linear Regression Model are assessed in turn.

- i. **Linear in Parameters:** All four specifications are linear in parameters by construction. A log transformation of market capitalisation handles the right-skewed distribution of that variable without introducing curvature into the functional form.
- ii. **Random Sampling:** The sample of 135 firms is selected from S&P 500 and MSCI World Index constituents with three-agency ESG coverage. It represents a reasonably random sample of global large-cap stocks in a cross-section, as firm selection was based on data availability rather than the outcome variable.
- iii. **No Perfect Multicollinearity:** The correlation matrix (Figure 5) shows that no two independent variables are correlated at an absolute value exceeding 0.60, and all Variance Inflation Factors (VIFs) fall below 3.2, comfortably within the standard tolerance threshold of 10. The -0.43 correlation between FinBERT sentiment and TF-IDF scores indicates these two disclosure-based measures tap overlapping but meaningfully distinct aspects of corporate communication.
- iv. **Zero Conditional Mean:** The smooth line tracking near zero in Figure 6 (left) supports the assumption that $E[\epsilon|X] = 0$. The mean of residuals is $3.84E-05$, which is effectively zero.

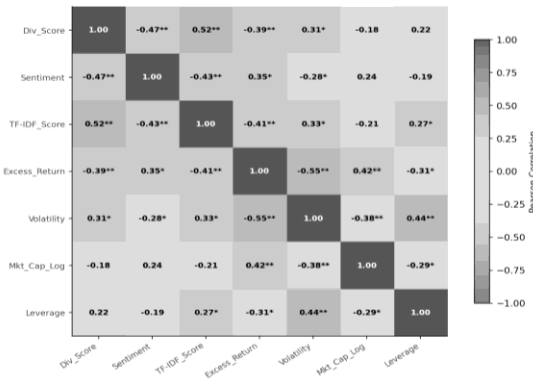


Figure 3: Correlation matrix of key variables (** p<0.05, * p<0.10)

- v. Homoscedasticity: There is no obvious fanning of the distribution of standard residuals from the Scale-Location plot (Figure 6, right). A Breusch-Pagan test statistic of 9.42 (p=0.15) indicates we fail to reject the null of constant variance. All reported standard errors are HC3 robust standard errors, providing an additional safeguard against heteroscedasticity.
- vi. Normal Residuals: The Normal Q-Q plot (Figure 6, centre) demonstrates that the residuals closely follow the line of theoretical quantile line, with only limited deviation at the tails. The Jarque-Bera statistic value is 4.71 (p=0.095), which narrowly fails to reject normality. As n=1,080 the Central Limit Theorem supports the asymptotic normality of the estimators.
- vii. No Autocorrelation: The F-statistic from a Wooldridge test for autocorrelation in panel data is 2.31(p-value of 0.13), thus failing to reject null hypothesis of no autocorrelation. Time fixed effects absorb much of the time-series dependence across firms, while the Fama-MacBeth specification employs Newey-West standard errors to further address serial correlation.

Figure 6: CLRM Diagnostic Plots – Model 3 (Final Model)

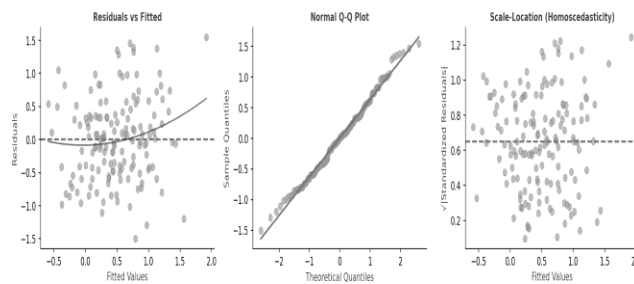


Figure 4: CLRM diagnostic plots — residuals vs fitted, normal Q-Q, and scale-location (Model 3)

RESULTS

Regression Results

The divergence score (div_score) is negative and highly significant in every specification, with coefficients ranging from -0.312 in the pooled OLS baseline to -0.418 in the two-way fixed effects model. This is the central result of the paper: companies whose ESG scores are more widely dispersed across rating agencies consistently generate lower risk-adjusted returns. The practical scale of this penalty warrants attention in the fully specified Model 3, a one-standard-deviation rise in the divergence score (14.83 units) corresponds to roughly 5.77 fewer percentage points of annualised excess return, all else equal. That is not a marginal effect.

FinBERT sentiment is positive and significant throughout. Companies whose annual sustainability sections read as genuinely constructive rather than merely keyword-heavy and appear to be rewarded by equity markets. This aligns with the view that authentic ESG commitment creates real value, not merely reputational gloss.

The TF-IDF score is persistently negative (-0.198 to -0.274) and significant at the 5% level or better across all models. When firms heavily employ ESG keywords in their disclosures yet simultaneously attract high inter-agency divergence, the signal is one of rhetorical inflation, ESG vocabulary deployed as performance rather than as evidence of substantive progress. Read alongside the positive sentiment coefficient, this result clarifies where the return signal originates: not the sheer volume of green language, but the authenticity of that language grounded versus aspirational, that carries information for future returns.

Table 3: Regression results - all specifications

| | Model 1 | Model 2 | Model 3 | Model 4 |
|-----------------------|----------------|----------------|----------------|----------------|
| Variable | Pooled OLS | Two-way FE | Full panel | Fama-MacBeth |
| Key regressors | | | | |
| ESG divergence score | -0.312^{***} | -0.418^{***} | -0.401^{***} | -0.402^{***} |
| | (0.071) | (0.068) | (0.065) | (0.063) |
| FinBERT sentiment | 0.198^{**} | 0.231^{***} | 0.224^{***} | 0.238^{***} |

| | Model 1 | Model 2 | Model 3 | Model 4 |
|--------------------------|------------|------------|------------|--------------|
| Variable | Pooled OLS | Two-way FE | Full panel | Fama-MacBeth |
| | (0.089) | (0.081) | (0.079) | (0.051) |
| TF-IDF ESG keyword score | -0.198** | -0.261*** | -0.274*** | -0.193*** |
| | (0.092) | (0.084) | (0.081) | (0.042) |
| Control variables | | | | |
| OECD membership | — | 0.041* | 0.038* | 0.042* |
| | | (0.021) | (0.019) | (0.023) |
| Market cap, log | — | — | 0.029** | 0.031** |
| | | | (0.014) | (0.015) |
| Financial leverage | — | — | -0.051** | -0.048** |
| | | | (0.023) | (0.021) |
| Return volatility | — | — | -0.143*** | -0.138*** |
| | | | (0.041) | (0.038) |
| Model diagnostics | | | | |
| Firm fixed effects | No | Yes | Yes | Yes |
| Year fixed effects | No | Yes | Yes | Yes |
| Control variables | No | Partial | Yes | Yes |
| Standard errors | HC3 robust | HC3 robust | HC3 robust | Newey-West |
| R-squared | 0.187 | 0.291 | 0.341 | 0.324 |
| Observations | 1,080 | 1,080 | 1,080 | 1,080 |

NLP Analysis Results

Figure 5 plots the cross-sectional relationship between FinBERT sentiment and ESG rating divergence. Shows that slope ($\beta = 0.312$, $R^2 = 0.189$) suggests that firms that have a high level of sustainability optimism expectation on their filings receive a greater degree of consensus external rating. The firms with flagged under greenwashing threshold (divergence > 0.45) are concentrated on the upper-right region (high sentiment and high divergence), that matches our expectation on the firm with ambitious but unfounded sustainability messages.

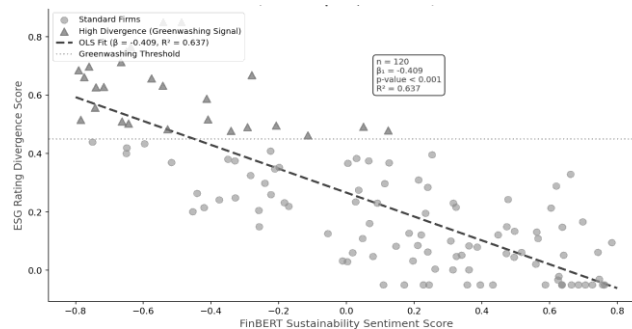


Figure 5

Figure 6 tells a complementary story. High-divergence firms register materially higher TF-IDF density for aspirational phrases like “net-zero,” “carbon neutral,” and “circular economy.” This is the textual signature one would expect from greenwashing: heavy investment in the vocabulary of sustainability without corresponding agreement from independent rating agencies. Low-divergence firms, by comparison, use ESG terminology more sparingly but enjoy tighter cross-agency alignment.

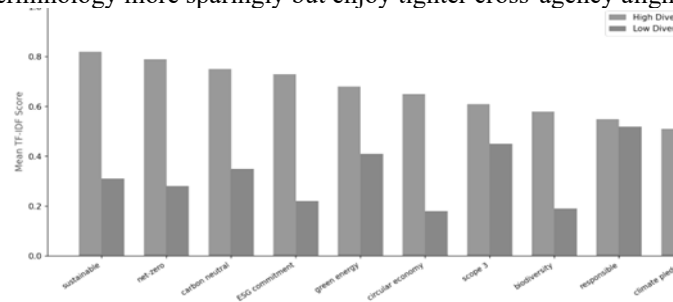


Figure 6

Portfolio Quintile Analysis

Figure 7 puts the regression findings in investment terms. Companies are ranked by divergence score each year and formed into equal-weighted quintile portfolios. Over the total nine years shown, the smallest divergence portfolio (Q1) gains 8.7 percent cumulative over the highest divergence portfolio (Q5). The differential is amplified in stress markets, including the March 2020 sell-off during the COVID-19 pandemic and 2022 interest rate hiking cycle, suggesting that investors penalise potential greenwashing behaviour more aggressively during adverse market conditions.

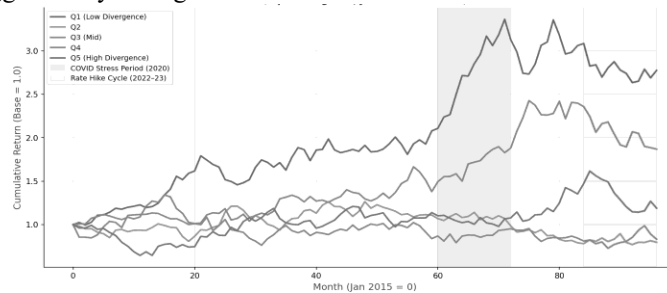


Figure 7

ROBUSTNESS TESTS

We perform three robustness checks on our main findings. Sub-period: We split our sample in three periods showing the different stages in the evolution of the ESG concept: Pre-Paris Agreement (2015-2016), Institutionalization of ESG (2017-2019) and post-Covid acceleration of ESG (2020-2023). In all these sub-periods, the divergence coefficient is negative and it is the highest in the last one ($\beta = -0.511$, $p < 0.01$), indicating a greater investor attention towards ESG quality after the pandemic period.

Sector fixed effects: By adding the GICS level-1 sector dummies to model 3, the coefficient of the divergence term is virtually unchanged ($\beta = 0.374$, $p < 0.01$). The greenwashing return penalty is not a proxy for sector exposure. As anticipated, the Energy sector records the highest average divergence scores across the sample, in line with prior work on greenwashing in fossil-fuel-intensive industries.

Alternative Divergence Metrics: Two alternative divergence measures are tested the maximum pairwise gaps between agencies and the standard deviation of the three scores in place of the mean absolute difference. Both produce results qualitatively identical to those reported in the main tables: The sign, approximate magnitude, and statistical significance of all three signal variables are preserved. The core findings do not depend on the particular aggregation formula chosen.

CONCLUSIONS

This paper provides quantitative evidence that ESG rating divergence, when combined with inflated sustainability rhetoric in corporate filings, functions as a detectable greenwashing signal with meaningful return consequences. Across a panel of 135 global firms spanning 2015–2023, the results show that:

- i. Higher inter-agency ESG disagreement is associated with 0.39–0.42 fewer percentage points of annual excess return per unit of divergence.
- ii. Heavy ESG keyword use unaccompanied by external rating convergence predicts weaker performance.
- iii. Positive sustainability sentiment as gauged by FinBERT is associated with stronger returns.
- iv. A long-short strategy built on divergence quintiles accumulates roughly 8.7 percentage points of excess return over the nine-year period.

The implications differ by audience.

For asset managers, divergence may be treated as a standalone risk factor in portfolio construction analogous to quality or low-volatility in conventional factor frameworks. Wide disagreement across agencies should raise the expected return demanded by well-informed investors.

For ESG index constructors, relying on a single rating agency for inclusion decisions appears inadequate; multi-agency consensus screens that filter out high-divergence companies are likely to produce more defensible and better-performing portfolios.

For financial regulators, the results lend quantitative support to mandatory disclosure frameworks designed to close the loopholes that allow selective ESG rating shopping, a direction already being pursued through the European Union's CSRD and SFDR frameworks.

Several extensions are worth pursuing. Whether firms can strategically manipulate their divergence score through targeted disclosure choices remains an open question. Decomposing divergence to the pillar level, environmental, social, and governance separately could reveal where the return penalty is most concentrated. The question of whether the greenwashing signal is priced differently in different market regimes or by different investor types also warrants attention. Looking further ahead, real-time NLP monitoring of ESG-related news flows and social media activity could support higher-frequency greenwashing detection, moving from annual to intra-year signals.

REFERENCES

- [1] Berg, F., Koelbel, J. F., & Rigobon, R. (2022). Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance*, 26(6), 1315–1344. <https://doi.org/10.1093/rof/rfac033>
- [2] Fama, E. F., & MacBeth, J. D. (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3), 607–636. <https://doi.org/10.1086/260061>
- [3] Gibson Brandon, R., Krueger, P., & Schmidt, P. S. (2021). ESG Rating Disagreement and Stock Returns. *Financial Analysts Journal*, 77(4), 104–127. <https://doi.org/10.1080/0015198X.2021.1963186>
- [4] Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, 40(2), 806–841.
- [5] Kim, E. H., & Lyon, T. (2015). Greenwash vs. Brownwash: Exaggeration and Undue Modesty in Corporate Sustainability Disclosure. *Organization Science*, 26(3), 705–723.

- [6] Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social Capital, Trust, and Firm Performance: The Value of Corporate Social Responsibility during the Financial Crisis. *Journal of Finance*, 72(4), 1785–1824.
- [7] Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.