

Article

Carbon Risk Without a Stable Premium: Nonlinear and State-Dependent Evidence from European ESG Leaders

Eleonora Salzmann 

Department of Finance, Faculty of Economic Studies, University of Finance and Administration, Estonská 500, 101 00 Prague, Czech Republic; 39401@mail.vsfs.cz or ae.salzmann@gmail.com; Tel.: +49-173-802-91-29

Abstract

Despite the economic relevance of climate-transition risk, firm-level carbon exposure often fails to appear as a robustly priced factor when ESG measures and sustainability shocks are conflated. This study examines whether carbon exposure is conditionally priced in European equity returns using a strongly balanced quarterly panel of 238 firms from the MSCI Europe ESG Leaders universe (2018–2024). Total greenhouse gas emissions act as a proxy for carbon exposure, mapped to within-year percentiles and standardized by sector-year. Regressions control for ESG scores and controversies and include firm and quarter fixed effects with firm-clustered, dependence-robust standard errors. The linear carbon coefficient is small and statistically indistinguishable from zero, indicating no stable return premium from within-firm changes in carbon exposure. Functional-form tests reject linearity: quadratic and quintile specifications reveal curvature and a non-monotonic pattern, with return differences concentrated in the middle of the carbon distribution. Conditioning on macro-financial stress, measured by the ECB Composite Indicator of Systemic Stress, yields limited evidence of a uniform carbon penalty. However, high-controversy states are associated with lower returns, while ESG scores show negative associations under dependence-robust inference. Overall, carbon-related pricing appears to be nonlinear and state-dependent, whereas controversy risk is the most robust sustainability predictor of returns.

Keywords: ESG factors; asset pricing; European equity markets; carbon transition risk; ESG controversy risk

1. Introduction

The transition to a low-carbon economy has elevated firm-level emissions from a mere disclosure item to a potential pricing input. Carbon-intensive business models face transition-related threats through multiple channels, including regulatory tightening, technological substitution, and shifts in investor demand. Yet empirical evidence on asset pricing remains unsettled (Bolton and Kacperczyk 2021, 2023; Ehlers et al. 2021; Pástor et al. 2021). Some studies link return differentials to emissions exposure, whereas others emphasize that climate-transition risk may manifest less as a stable means-return premium and more through downside, crash, and financing channels (Ilhan et al. 2021; Bose et al. 2025). This tension raises a fundamental question for both academics and practitioners: If carbon exposure is financially material, why does it often fail to appear as a robustly priced factor in standard return regressions?

A central challenge is that sustainability information is multi-dimensional and arrives in diverse forms. Carbon exposure is a structural characteristic associated with transition



Academic Editor: Alejandro Balbás

Received: 20 January 2026

Revised: 6 February 2026

Accepted: 12 February 2026

Published: 20 February 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

risk, but it is frequently proxied by imperfect and evolving emissions data. At the same time, the sustainability signals most commonly used by investors (composite ESG scores) are noisy, provider-dependent, and only indirectly related to environmental exposure (Bolton and Kacperczyk 2021, 2023; Berg et al. 2022; Serafeim and Yoon 2022). Moreover, sustainability-related downside risk can be triggered by discrete controversy events, such as incidents, litigation, or regulatory breaches, which resemble shocks rather than slow-moving characteristics (Serafeim and Yoon 2022; Bang et al. 2023). These three components (structural carbon exposure, noisy ESG composites, and event-driven controversies) are conceptually distinct and may influence returns through different mechanisms. Treating them as interchangeable, or compressing them into a single “ESG factor,” risks obscuring relevant pricing variation.

This paper isolates these dimensions and addresses the following question: Is firm-level carbon exposure conditionally priced in European equity returns once broad ESG characteristics and controversy states are controlled for? We formalize this inquiry in H1 (Conditional Carbon Pricing Hypothesis): firm-level carbon exposure is not associated with stock returns on average; however, it becomes negatively priced during periods when carbon-related risks attain economic significance. The hypothesis posits that transition risk is state-dependent: carbon exposure may not command a constant premium under normal conditions but may become more relevant when uncertainty rises and risk-bearing capacity is constrained.

Empirically, the analysis relies on a balanced quarterly panel designed to mitigate common pitfalls in ESG research, including missing data, inconsistent measurement, and shifting sample composition. The dataset comprises 238 European firms from the MSCI Europe ESG Leaders universe over 2018–2024, yielding 6664 firm-quarter observations. Returns are sourced from S&P Capital IQ Pro, sustainability and controversy information is sourced from professional ESG data providers, and firm-level carbon exposure is proxied by total reported greenhouse gas emissions, consistently aligned over time. To capture macro-financial conditions coinciding with periods of economically significant carbon risk, the study employs the ECB Composite Indicator of Systemic Stress (CISS) as a quarter-aligned, time-varying stress proxy.

The econometric design prioritizes conservative identification. The baseline specification is a two-way fixed-effects regression with firm and quarter fixed effects, absorbing time-invariant firm heterogeneity (e.g., persistent business-model differences) and common shocks within each quarter. This within-firm approach is essential because many carbon and ESG characteristics are persistent and strongly correlated with stable firm attributes; failing to control for these factors risks conflating cross-sectional “firm type” differences with dynamic pricing effects. To address functional-form uncertainty and avoid imposing a linear carbon premium, the analysis complements the baseline specification with quadratic and nonparametric (quintile) representations of carbon exposure. To operationalize conditional pricing, the framework further incorporates carbon–stress interaction terms. Inference is validated using dependence-robust approaches suitable for financial panels that may exhibit heteroskedasticity, within-firm autocorrelation, and cross-sectional dependence.

The contribution of this paper is three-fold. First, it provides a clean within-firm European panel test of conditional carbon pricing in a benchmark-driven, investable universe with strong data coverage, reducing the risk of results being driven by sample churn or missing ESG histories. Second, it treats carbon exposure, ESG composites, and controversy states as separate factors in the regression design, aligning the empirical specification with conceptual distinctions emphasized in the literature. Third, it embeds state dependence and functional-form flexibility directly into the empirical framework and

evaluates results under dependence-robust inference, which is essential when sustainability variables are persistent and return data are heavy-tailed.

The key empirical novelty of this study is that it (i) treats structural carbon exposure (emissions), composite ESG assessments, and event-driven controversy states as separate sustainability dimensions rather than interchangeable proxies; (ii) allows the carbon–return relationship to be nonlinear (quadratic and nonparametric carbon-bucket specifications) instead of imposing a constant “brown-minus-green” slope; and (iii) tests for state dependence by examining whether carbon sensitivity changes when market-wide stress or transition salience rises. This design speaks directly to why carbon premia can appear unstable in standard regressions: combining emissions with noisy ESG composites and controversy shocks—and forcing linearity—can average out localized, threshold-like, and regime-dependent pricing patterns.

The remainder of the paper is structured as follows. Section 2 reviews the theoretical background and the empirical literature on carbon risk, ESG measurement, and controversy risk. Section 3 describes the data, variables, and construction of carbon exposure and sustainability controls. Section 4 outlines the econometric methodology, including fixed-effects identification, stress conditioning, functional-form checks, and robust inference. Section 5 presents empirical results. Section 6 discusses the findings in relation to the existing literature and the Conditional Carbon Pricing Hypothesis. Section 7 presents the conclusion and highlights limitations and directions for future research.

2. Theoretical Background and Related Literature

This section reviews three strands of the literature that inform the empirical design. First, it summarizes evidence on carbon exposure as a financially relevant risk factor in asset pricing. Second, it discusses ESG characteristics, emphasizing measurement noise and the weak pricing performance of composite ESG scores. Third, it examines ESG controversies as short-horizon, event-driven sources of downside risk. Together, these strands motivate an empirical framework that isolates the conditional pricing of firm-level carbon exposure while controlling for broader sustainability signals and episodic controversy effects.

2.1. Carbon Risk and Asset Pricing

Carbon exposure¹ has emerged as a financially material source of risk in equity markets, reflecting firms’ vulnerability to the transition toward a low-carbon economy. Unlike composite ESG scores, carbon-related metrics² are directly linked to transition risks arising from regulatory changes, technological substitution, and shifts in investor demand away from carbon-intensive business models (Bolton and Kacperczyk 2021, 2023). Empirical evidence associates higher emissions with systematically different stock returns, although the magnitude and direction of estimated premia vary across markets and over time (Bolton and Kacperczyk 2021, 2023; Ilhan et al. 2021; Bose et al. 2025). Recent studies also document carbon-transition-risk premia in additional markets, including China, and link carbon-risk exposure to broader corporate policies such as payout decisions (Luo and Ma 2025; Zhou et al. 2025; Boubaker et al. 2024).

Global evidence indicates that carbon-transition premia vary across regions and regulatory regimes. Carbon premia demonstrate economic significance in certain markets and periods, such as Europe and Asia following the Paris Agreement, but are less pronounced in other contexts. This variation supports the perspective that pricing is influenced by domestic climate-policy stringency and investor awareness, rather than by a uniform global premium (Bolton and Kacperczyk 2023).

Average return differentials do not fully capture carbon risk, which instead manifests through downside and tail risk. Ilhan et al. (2021) show that carbon-intensive firms face

higher costs of protection against extreme negative events, suggesting that carbon risk is reflected in crash and volatility risk rather than stable return premia. Consistent with this view, Bose et al. (2025) provide international evidence that elevated carbon risk predicts higher future stock price crash risk, particularly in environments characterized by greater information asymmetry.

Other studies highlight valuation and risk channels beyond unconditional return effects. Al Rabab'a et al. (2024) found that stronger corporate carbon performance is associated with lower total, systematic, and idiosyncratic risk. Duppati (2025) shows that increases in carbon pricing reduce firms' revenue growth, thereby raising valuation risk via a cash-flow channel. Credit market evidence also indicates higher borrowing costs for carbon-intensive firms, implying that carbon risk is incorporated into financing conditions even when equity-return premia are weak or unstable (Ehlers et al. 2021; Dong et al. 2025). Overall, these findings indicate that carbon risk affects firm value through both discount-rate and cash-flow channels. This heterogeneity also helps reconcile conflicting findings in the return-based literature. When carbon risk is primarily priced through tail risk or financing conditions rather than average returns, conventional return regressions may yield unstable or insignificant premia, even if carbon exposure is economically significant. Consistent with this interpretation, evidence from syndicated loans demonstrates a significant carbon premium since the Paris Agreement. This finding suggests that transition risk can be internalized through financing costs, even when equity premia are weak or episodic.

Collectively, the literature implies state-dependent pricing of carbon exposure. When investors pay limited attention to carbon-related risks, estimated return effects may appear weak or unstable. When uncertainty rises (due to heightened investor attention or transition-related shocks), carbon exposure becomes more relevant for pricing through downside-risk channels (Ilhan et al. 2021; Bose et al. 2025). This motivates empirical approaches that allow the return–carbon relationship to vary over time rather than imposing a constant carbon premium.

2.2. ESG Characteristics and Measurement Noise

ESG scores are widely used as broad indicators of corporate sustainability; however, their interpretation as asset-pricing factors remains contested. ESG ratings exhibit substantial disagreement across providers due to differences in data sources, weighting schemes, and aggregation methodologies (Berg et al. 2022). These discrepancies are particularly pronounced within the environmental pillar, where ratings integrate emissions data with qualitative evaluations of corporate policies, disclosure practices, and governance structures. Consequently, ESG scores provide an imprecise and noisy measure of firms' underlying environmental exposure.³ Evidence also indicates that carbon accounting systems and disclosure infrastructure can shape observed carbon performance and its link to strategy, reinforcing the importance of measurement choices for empirical inference (Bui et al. 2022).

Consistent with this concern, ESG disagreement weakens the relationship between ESG scores and realized environmental outcomes. Deng et al. (2025) find that greater divergence across ESG ratings is associated with higher carbon emission intensity. Studies focusing on ESG disclosure rather than ESG scores document heterogeneous effects on carbon intensity that depend on firm characteristics such as profitability, intangible asset intensity, and innovation capacity (Saha and Maji 2025; Xie et al. 2024). Complementary evidence, particularly from China, reports carbon-intensity reductions associated with ESG performance across sectors, including listed manufacturing and shipping firms, and highlights that this relationship is shaped by climate-policy uncertainty and digitalization (Li et al. 2024; Li et al. 2025; Ye and Xu 2023).

Asset-pricing evidence indicates weak and unstable return effects associated with ESG characteristics, suggesting that broad ESG signals are largely diversified at the market level (Pástor et al. 2021; Monasterolo and de Angelis 2020). Equilibrium models similarly imply that ESG-related return differentials can reflect shifts in investor preferences rather than compensation for systematic risk (Pástor et al. 2021). Accordingly, ESG variables are best treated as controls that capture broad sustainability and reputational signals, allowing analysis to isolate the pricing of more directly measurable risk exposures such as carbon emissions.

Overall, ESG scores are not direct measures of carbon risk; they provide broad and noisy sustainability signals that only partially reflect firms' environmental exposure. Therefore, we treat ESG variables as controls capturing general sustainability and reputational information while isolating the pricing effects of directly measurable carbon exposure.

2.3. Controversy Risk as Downside Exposure

In addition to slow-moving ESG characteristics and structural carbon exposure, a distinct strand of the literature focuses on ESG controversies, which capture event-driven downside risk arising from sudden adverse information about firms' environmental, social, or governance practices (Bang et al. 2023; Serafeim and Yoon 2022). Controversy indicators reflect discrete events, such as environmental accidents, regulatory violations, litigation, or governance scandals, which trigger abrupt reassessments of firm value. Thus, controversy risk is inherently short-horizon and asymmetric.

The empirical literature suggests that controversy-related effects are primarily priced through downside and tail-risk channels rather than average-return effects. Ilhan et al. (2021) document heightened tail-risk pricing for carbon-intensive firms, and Bose et al. (2025) show that carbon-related crash risk persists even after controlling for ESG scores, indicating that composite ESG ratings do not fully capture sustainability-related downside risk. Bang et al. (2023) further demonstrate that ESG controversies represent event-driven risk components distinct from average ESG characteristics and command a separate risk premium. These findings imply that controversy indicators proxy episodic shocks rather than persistent firm characteristics.

From an asset-pricing perspective, controversy risk represents a non-diversifiable, event-driven component that can affect returns independently of structural carbon exposure. Controlling for controversy indicators prevents attributing event-driven losses to slow-moving sustainability characteristics and improves the identification of carbon-exposure pricing effects. Accordingly, the analysis includes ESG controversy indicators as controls for short-horizon downside risk, separating structural carbon exposure from event-driven sustainability shocks in the return–carbon relationship.

2.4. Summary and Empirical Implications

The three strands reviewed above indicate that carbon exposure, ESG characteristics, and ESG controversies constitute distinct dimensions of sustainability-related risk. Carbon exposure reflects structural transition risk that becomes economically significant under certain market conditions. ESG scores provide broad but imprecise sustainability signals that are weakly incorporated into prices. ESG controversies represent short-term, event-driven downside risk. These dimensions can interact empirically. ESG scores may incorporate some emissions-related information, but this is often accompanied by significant measurement noise. In contrast, controversies can introduce abrupt downside shocks that dominate short-horizon returns. If structural carbon exposure is not separated from noisy composite measures and event-driven shocks, the observed relationship between carbon and returns may appear weak or unstable, even when carbon risk is conditionally priced. Importantly,

both theory and empirical evidence suggest that such pricing effects are unlikely to manifest as a stable, linear return premium. Instead, carbon exposure may become return-relevant only when transition risk attains economic salience—such as during periods of heightened uncertainty, policy attention, or constrained risk-bearing capacity—implying nonlinear and state-dependent pricing rather than a constant carbon premium.

Based on this distinction, the empirical analysis designates firm-level carbon exposure as the primary explanatory variable in return regressions, while ESG scores and ESG controversy indicators serve as controls. This approach isolates the pricing of structural carbon exposure from composite sustainability characteristics and episodic controversy effects and permits the return–carbon relationship to vary over time. Accordingly, the study tests the following hypothesis (Conditional Carbon Pricing Hypothesis):

H1. *Firm-level carbon exposure is not associated with stock returns on average; however, it becomes negatively priced during periods when carbon-related risks become economically significant.*

3. Data and Variables

3.1. Data Sample of MSCI Europe ESG Leaders

The empirical analysis uses quarterly data for European-listed firms from 2015Q1 to 2024Q4, with the estimation sample covering 2018Q1 to 2024Q4. Firms are drawn from the MSCI Europe ESG Leaders Index universe and included only if a consistent quarterly panel is available for returns, ESG pillar metrics, total greenhouse gas (GHG) emissions, and controversy indicators. Due to substantial gaps in ESG, emissions, and controversy data from 2015 to 2017, the primary analysis focuses on 2018 onward to maintain a balanced panel and support robust dynamic beta estimation and cross-sectional pricing tests. Applying these coverage and panel requirements yields a final balanced quarterly dataset of 238 firms⁴ from 2018 to 2024, comprising 6664 firm-quarter observations. Annual firm-level total GHG emissions (tCO₂e) are allocated to quarters using a consistent timing rule. Further details regarding emissions data sourcing and alignment are provided in Section 3.2.⁵

3.2. Dataset Description

The analysis integrates quarterly equity returns with ESG pillar metrics, controversy indicators, firm-level greenhouse gas emissions, and a macro-financial stress proxy for a balanced European panel. Table 1 summarizes the variables used, their frequency, and their data sources⁶.

Returns and benchmarks. Firm-level total returns, including dividends, and standard market and accounting identifiers are obtained from S&P Capital IQ Pro⁷, consistent with established empirical asset-pricing practice (Fama and French 1993; Hou et al. 2015). Market returns are proxied by the MSCI Europe Index, which tracks large- and mid-cap equities across 15 developed European markets and covers approximately 85% of the free-float-adjusted market capitalization of the European developed-markets equity universe. Excess returns are computed relative to the three-month EURIBOR money-market rate⁸ (ECB Data Portal/Statistical Data Warehouse) and aligned to the quarterly frequency, matching the quarterly return horizon.

ESG characteristics. The composite ESG score and E, S, and G pillar scores are sourced from S&P Global/MSCI ESG Research via Capital IQ. Using a professional provider framework enhances cross-firm comparability and aligns with the ESG asset-pricing literature that emphasizes standardized inputs and a clear link between sustainability characteristics and asset pricing (Pedersen et al. 2021).

Table 1. Description of variables used in the empirical model.

Variable	Description	Calculation Method/Source
Total Return	Firm-level quarterly total stock return, including dividends	Arithmetic total return from S&P Capital IQ Pro. When computed from levels: $R_{i,q} = \frac{TRI_{i,q}}{TRI_{i,q-1}} - 1$, where (TRI) is a total-return index (dividends reinvested).
Market Return	Quarterly European market benchmark return	Quarterly arithmetic return on MSCI Europe Index: $R_{m,q} = \frac{I_q}{I_{q-1}} - 1$, where the index return series is sourced via MSCI/Capital IQPro.
Risk-Free Rate	3-month risk-free rate (quarterly alignment)	3-month EURIBOR obtained from ECB Statistical Data Warehouse. Used to form excess returns where applicable; converted to a quarterly rate in decimals using a consistent convention (e.g., effective quarterly rate).
ESGScore (E, S, G)	Composite ESG score and E/S/G pillar scores (annual, aligned to quarters)	ESG factor returns constructed from ESG-sorted portfolios (as defined in the factor-construction step). Betas estimated via rolling-window regressions/Kalman filter on these factor returns. ESG inputs sourced from S&P Global/MSCI ESG Research (via Capital IQ Pro).
ESGControversy Level	ESG controversy level/severity (quarterly)	Portfolio returns sorted by level of controversy; Systanalytics, Corporate and Investor reporting.
Carbon Score	Relative firm carbon exposure based on total GHG emissions (annual, aligned to quarters)	Assigned based on reported carbon emissions (total GHG) provided by DitchCarbon and used as an explanatory factor in second-stage panel regressions; not used in portfolio sorting. CarbonScore constructed via within-year cross-sectional normalization: $CS_{i,y} = \frac{GHG_{i,y} - \min_j GHG_{j,y}}{\max_j GHG_{j,y} - \min_j GHG_{j,y}}$, then mapped to quarters using a consistent carry-forward rule.
CISS	Macro-financial systemic stress control (quarterly)	ECB Composite Indicator of Systemic Stress (ECB Statistical Data Warehouse); aligned to quarters and used as macro control (optionally standardized).
EUA	Carbon-price-shock proxy (quarterly)	EU ETS allowance price shock ($\Delta \ln(\text{EUA})$, quarterly; standardized) reflects direct European transition-policy cost signal.
TRI	Transition Risk Index (quarterly)	News-based transition risk index (quarterly average; standardized).
Country Dummy	Systemic stress proxy (quarterly)	Indicator set for firm domicile (15 MSCI DM Europe markets); one omitted reference category.
Quarter Dummy	Country fixed effects (time-invariant firm domicile)	Indicator set for each calendar quarter in the sample (e.g., 2018Q1–2024Q4); one omitted reference quarter.

Source: Author calculations based on data from S&P Capital IQ Pro, Sustanalytics, DitchCarbon, ECB Statistical Data Warehouse, and market sources, as detailed in the table. **Note:** This table describes the construction, frequency, and data sources of all variables used in the empirical model. Firm-level variables are aligned to quarterly frequency. Carbon exposure is based on total reported GHG emissions and cross-sectional normalization. Transition-risk proxies (EUA, TRI) and CISS are standardized when used in regressions.

ESG controversies. Controversy levels are obtained from Sustainalytics⁹. For periods with incomplete historical coverage, controversy histories are extended using hand-collected events from public firm disclosures, mapped to the Sustainalytics framework. This approach follows the view that controversies represent reputational-risk shocks distinct from long-term ESG characteristics (Ilhan et al. 2021; Bang et al. 2023).

Carbon exposure. Carbon exposure is measured using total GHG emissions (tCO₂e). The primary source is DitchCarbon¹⁰, which provides coverage for 98 firms; for the remaining firms, total GHG values are manually collected from corporate sustainability reports and regulatory disclosures using a uniform rule (consolidated total GHG, converted to tCO₂e, aligned to the reporting year). A single total GHG measure is used rather than aggregating Scope 1–3 components because scope boundaries and coverage are often incomplete and inconsistently defined across firms and years, particularly in earlier periods. This approach is consistent with the carbon-risk asset-pricing literature that treats emissions exposure as a distinct transition-risk channel (Bolton and Kacperczyk 2021; Ilhan et al. 2021; Aswani et al. 2024).¹¹

Macro-financial stress control. Systemic stress is controlled using the ECB CISS from the ECB Statistical Data Warehouse (see Note 8 above). CISS is a composite measure designed to capture system-wide financial stress rather than isolated volatility and is based on the framework developed by Holló et al. (2012). Including CISS is motivated by evidence that risk premia and ESG-related pricing effects are regime-dependent and may vary under stress conditions.

Transition-risk conditioning variables.¹² Although CISS captures broad changes in macro-financial stress and aggregate risk-bearing capacity, it is not designed to isolate carbon transition shocks. To align the conditioning information set more closely with transition risk, we therefore complement CISS with two transition-specific variables constructed at a quarterly frequency. First, we use EU ETS allowance price shocks (EUA)¹³ as a market-based measure of variation in the marginal cost of emissions under the European cap-and-trade regime; quarterly changes in allowance prices provide a direct signal of tightening (or easing) transition cost pressure (Ehlers et al. 2021). Second, we use a news-based Transition Risk Index (TRI)¹⁴ aggregated to a quarterly frequency, which captures time variation in transition-risk-related information in the public news environment (Bua et al. 2022). Because both variables vary only over time, their main effects are absorbed by quarter fixed effects; identification therefore relies on interaction terms with firm-level carbon exposure (and carbon quintiles), consistent with a conditional pricing framework.

Country and quarter dummies (fixed effects). To account for cross-country heterogeneity in disclosure regimes, institutional environments, and market structure, the baseline specifications include country fixed effects based on firms' domicile countries. These dummies absorb time-invariant differences across jurisdictions, ensuring that identification relies on within-country variation. Quarter fixed effects (qdate dummies) are also included to control for common time shocks affecting all firms in a given quarter, such as macroeconomic developments, regulatory announcements, and market-wide risk episodes. Country dummies are constructed as indicators for each domicile country (MSCI Developed Markets Europe classification), and quarter dummies are constructed as indicators for each calendar quarter in the estimation window (e.g., 2018Q1 to 2024Q4), with one category omitted as the reference to avoid multicollinearity. These dummy variables are not reported in Table 2 because descriptive statistics are not informative for fixed effects.

Table 2. Summary of descriptive statistics.

Variable	Observations (Firm–Quarter)	Mean	SD	Min	Med	Max
TotalReturn	6664	0.0282	0.1372	−0.3447	0.0262	0.4082
ESGScore	6664	56.4139	17.2635	14	54	91
EScore	6664	62.5398	20.0816	2	64	99
SScore	6664	53.0452	19.1028	9	49	98
GEScore	6664	55.3668	16.5378	13	53	92
ESGControversyLevel	6664	1.5206	0.6838	1	1	4
CarbonScore	6664	0.0182	0.0850	0	0.0140	1
MSCI Europe Index	6664	0.0089	0.0980	−0.2481	0.0212	0.1902
CISS	6664	0.5373	2.2923	−0.7152	−0.0323	11.5983
EUA	6664	0.0782	0.2018	−0.3319	0.0897	0.4846
TRI	6664	−0.0017	0.0038	−0.0080	−0.0028	0.0075
EURIBOR-3M	6664	−0.1345	1.0117	−5.2262	0.0009	1.0413

Source: Author calculations based on data from S&P Capital IQ Pro, Sustainalytics, DitchCarbon, ECB Statistical Data Warehouse, and market sources, as detailed in the table. Note: The table reports summary statistics for the balanced quarterly firm-level panel (2018Q1–2024Q4). Observations (firm–quarter) denote the number of firm–quarter observations. ESG scores are annual metrics aligned to quarterly frequency. CarbonScore reflects relative firm carbon exposure based on total GHG emissions and cross-sectional normalization. CISS, EUA, and TRI are quarterly macro/transition variables; EUA and TRI are standardized when used in regressions. EURIBOR-3M is expressed in annualized percentage terms.

4. Methodology

This section details the empirical workflow used to (i) assess the distributional and time-series properties of the data, (ii) construct risk-adjusted returns and comparable firm-level covariates without unnecessary observation loss, (iii) select a panel estimator consistent with the identification strategy, and (iv) validate inference using diagnostic tests and dependence-robust covariance estimators. Data sources and baseline variable definitions are provided in Section 3; the focus here is on econometric implementation and replication logic.

4.1. Diagnostics: Panel Structure, Descriptive Statistics, and Stationarity

Firm–time observations are organized as a quarterly firm-level panel spanning 2018Q1–2024Q4. The panel is strongly balanced, with each firm observed in every quarter, eliminating entry and exit concerns and ensuring that comparisons across estimators and robustness specifications are not affected by uneven time coverage. Prior to estimation, we assess the distributional properties, persistence, and stationarity of key variables to ensure that the empirical specifications are not driven by heavy tails or spurious time-series behavior. Given the slow-moving nature of sustainability indicators, we implement transformations that preserve within-firm variation while maintaining economic interpretability. Carbon exposure is mapped onto an annual rank-based percentile measure, and controversies are encoded as discrete state and transition indicators. A detailed description of the diagnostic procedures and transformation definitions is provided in Appendix A.1.

4.2. Construction of Quarterly Excess Returns

The dependent variable is the firm’s quarterly excess return, defined as the quarterly total return minus the quarterly risk-free rate. The risk-free rate is obtained from a short-term money market rate reported in annualized percent terms and converted to a quarterly decimal rate using the corresponding scaling. Return construction is verified to avoid missingness or mechanical sample loss. Formal definitions are provided in Appendix A.2.

4.3. Normalization and Outlier Handling

To improve comparability across firms and time while preserving the balanced panel, we apply structured outlier handling and scaling procedures. Continuous variables are winsorized at the 1st and 99th percentiles to reduce the leverage of extreme tail realizations without deleting observations, which is particularly relevant in stress-oriented return models. Key firm-level characteristics are then standardized within sector–time cells to remove mechanical industry-level differences and express covariates in comparable units. Where Sector $\times$ Quarter cells are thin, we use a less granular Sector $\times$ Year standardization in the core regressions to preserve sector neutrality while minimizing missingness. Time-varying state variables that enter interaction terms (e.g., systemic stress and transition proxies) are standardized using global z scores to support numerical stability and “per one standard deviation” interpretation. Formal definitions are provided in Appendix A.3.

4.4. Model Selection and Correlated Random Effects

Estimator choice follows established panel econometric logic, guided by formal specification tests and identification considerations. We compare pooled ordinary least squares, random effects, fixed effects, and between-effects specifications, including quarter effects. The goal is consistent estimation of within-firm associations between excess returns and carbon exposure while accounting for unobserved, time-invariant firm heterogeneity.

Panel effects are assessed using the Breusch–Pagan Lagrange Multiplier test for random effects, which evaluates whether the firm-specific error component is nonzero relative to pooled OLS. The relative consistency of random versus fixed effects is evaluated using the Hausman specification test, which examines whether regressors are orthogonal to unobserved firm heterogeneity. Given the economic plausibility that carbon exposure and sustainability characteristics correlate with persistent firm attributes (e.g., business model, technology, and disclosure practices), fixed effects are retained as the conservative baseline. Recognizing that Hausman statistics may be unstable in specifications with extensive time effects and near-collinearity, estimator choice is grounded in a combination of tests and economic identification logic rather than a single statistic.

Between-effects estimates are reported as a descriptive cross-sectional benchmark based on time averages but are not used for primary inference because they do not exploit within-firm variation, which is central to identifying carbon exposure and stress-sensitivity effects.

To directly test the core exogeneity restriction implied by random effects, we implement a correlated random-effects diagnostic using the Mundlak–Chamberlain device by augmenting the random-effects specification with firm-level means of key time-varying regressors. Statistical significance of these firm-mean terms indicates correlation between regressors and the unobserved firm effect, reinforcing the fixed-effects approach and providing a transparent decomposition of within- and between-firm components.

4.5. Empirical Specifications, Stress Conditioning, and Robustness Checks

All core return regressions are estimated using firm fixed effects (α_i) and quarter fixed effects (δ_t). Our estimand is within-firm repricing of changes in reported emissions within an ESG-screened investable universe; we do not claim to estimate population-wide cross-sectional carbon premia. The baseline specification is:

$$r_{i,t}^{ex} = \alpha_i + \delta_t + \beta Carbon_{i,t} + \theta Controv_{i,t} + \phi ESG_{i,t} + \varepsilon_{i,t},$$

where $Carbon_{i,t}$ denotes the sector-standardized carbon-exposure measure and the control set includes controversy indicators and standardized sustainability measures.

To examine whether carbon risk is priced differently under stressed market conditions, we extend the model with a stress-conditioning interaction:

$$r_{i,t}^{ex} = \alpha_i + \delta_t + \beta Carbon_{i,t} + \gamma (Carbon_{i,t} \times Stress_t) + \theta Controv_{i,t} + \phi ESG_{i,t} + \varepsilon_{i,t},$$

where $Stress_t$ is measured by a systemic stress index. Because the stress index varies only over time, its main effect is absorbed by the quarter fixed effects. In addition to CISS, we estimate parallel conditioning specifications using transition-specific state variables: EU ETS allowance price shocks and a news-based Transition Risk Index, each aligned to quarterly frequency and standardized. As with CISS, these variables vary only over time and therefore enter the model exclusively through interactions with firm-level carbon exposure (and carbon quintiles) under quarter fixed effects. Identification, therefore, relies on the interaction term, capturing whether cross-sectional carbon exposure is differentially priced during transition and stressed periods.

As an additional identification check, we estimate a specification that includes country indicators alongside quarter indicators. Since country affiliation is time-invariant at the firm level, country indicators are absorbed by firm fixed effects under the within estimator. This check confirms that the identifying variation in the empirical design arises from within-firm changes over time (including stress interactions), while country-level dependence is addressed through clustering (see Section 4.6).

Robustness checks address contemporaneous feedback, functional-form uncertainty, and tail sensitivity. First, to mitigate simultaneity concerns, we re-estimate the baseline model using one-quarter-lagged firm characteristics. Second, potential nonlinearity in carbon pricing is evaluated using a quadratic specification that includes both $Carbon_{i,t}$ and $Carbon_{i,t}^2$. Third, carbon exposure is discretized into quintiles to provide a nonparametric assessment of heterogeneity and monotonicity. We also estimate a central robustness model that interacts carbon quintile indicators with time-varying state variables—systemic stress (CISS) and transition-specific conditioning variables (EU ETS allowance price shocks and TRI)—allowing stress sensitivities to differ across the carbon distribution. The stress interaction block is evaluated using joint Wald tests, interpreted through marginal effects and contrasts relative to the low-carbon baseline, and visualized with predicted outcomes at representative stress levels. Finally, to ensure that results are not driven by extreme return observations, the dependent variable is winsorized at the 1st and 99th percentiles, and both baseline and stress-conditioning models are re-estimated under the same fixed-effects design.

4.6. Diagnostic Testing and Dependence-Robust Inference

Firm-level return panels typically exhibit heteroskedasticity, within-firm autocorrelation, and cross-sectional dependence arising from common shocks and market comovement. Accordingly, the empirical workflow combines model diagnostics with dependence-robust inference to ensure that statistical conclusions are not artifacts of restrictive error assumptions.

Multicollinearity is assessed using variance inflation factors to verify that carbon exposure, sustainability controls, and controversy indicators are not mechanically collinear. This check is particularly important in interaction models, as interaction terms can artificially inflate collinearity measures. Serial correlation is assessed using the Wooldridge test for autocorrelation in panel data, while groupwise heteroskedasticity is evaluated using the Modified Wald test across firms. Cross-sectional dependence is examined using complementary diagnostics, including Pesaran's CD test (and its absolute correlation variant), Friedman's test, and Frees' test, applied both to model residuals and raw specifications.

Examining residual dependence distinguishes shocks absorbed by quarter effects from dependence that persists beyond the model.

Inference is reported under a hierarchy of robust covariance estimators. Baseline results use firm-clustered standard errors, robust to heteroskedasticity and arbitrary within-firm serial correlation. Robustness checks include Driscoll–Kraay standard errors with a quarterly lag length of one year to address serial correlation and general cross-sectional dependence, two-way clustering by firm and quarter to allow dependence within firms and periods simultaneously, and country-level clustering in stress-conditioning models to account for national market co-movement and institutional shocks. In specifications with extensive time fixed effects, two-way clustered covariance estimators may become numerically unstable; in such cases we rely on firm-clustered and Driscoll–Kraay inference as the primary robustness checks. Results are considered robust when statistical significance and economic magnitudes remain stable across these increasingly conservative inference approaches.

4.7. Conceptual Challenges and Limitations

This study employs a reduced-form asset-pricing framework to assess whether firm-level carbon exposure is conditionally priced over time. Consistent with prior empirical research on carbon risk and sustainability factors, the estimated relationships represent conditional associations rather than structural causal effects attributable to a single mechanism (Bolton and Kacperczyk 2021, 2023; Pástor et al. 2021). Several limitations arise from this approach.

First, the principal sustainability variables are subject to measurement error. Carbon exposure is proxied by reported greenhouse gas emissions, which vary in coverage and reliability across firms and over time. ESG ratings incorporate provider-specific data choices and aggregation methodologies. Disagreement among ESG ratings indicates that these measures contain informational frictions and noise, potentially attenuating estimated coefficients and reducing comparability across samples (Berg et al. 2022; Deng et al. 2025). Moreover, evidence of heterogeneous ESG disclosure effects on carbon outcomes across firms suggests that ESG metrics are imperfect proxies for underlying environmental exposure (Saha and Maji 2025; Xie et al. 2024).

Second, the empirical analysis is constrained by data availability. Limiting the sample to firms with consistent emissions and ESG information enhances measurement credibility but may reduce external validity, as disclosing firms tend to be larger and more visible. Therefore, the findings should be interpreted as evidence regarding the pricing of observable carbon exposure among disclosing firms, rather than as population-wide estimates.

Third, multiple mechanisms may produce similar observed return patterns. Carbon-related return effects could result from risk compensation (Bolton and Kacperczyk 2021, 2023), downside-tail risk (Ilhan et al. 2021), or information-based channels related to crash risk and disclosure quality (Bose et al. 2025). Additionally, event-driven sustainability shocks may influence returns independently of persistent sustainability characteristics. ESG controversies represent distinct event-risk components that command separate premia (Bang et al. 2023). While the empirical design controls for ESG characteristics and controversy indicators to mitigate confounding, it does not disentangle the relative contributions of each underlying channel.

Fourth, selection and range-restriction considerations arise from the benchmark-driven sample design. Because the estimation sample is drawn from the MSCI Europe ESG Leaders universe and restricted to firms with consistent emissions coverage, the dispersion of carbon exposure may be truncated relative to the broader market (range restriction). This truncation can mechanically reduce the detectability of linear pricing effects, particularly in

within-firm fixed-effects specifications that rely on time variation, while leaving nonlinear or threshold-type patterns more visible. To address this concern transparently within the available data, the main estimations are complemented with (i) dispersion diagnostics for emissions and carbon proxies under the paper's transformations, (ii) within-between decompositions of carbon variation, and (iii) range-restriction sensitivity checks based on trimming extreme observations and firms. These diagnostics, reported in Appendix C, serve to delineate applicability boundaries rather than to alter the primary identification strategy.

Finally, conditional pricing results are sensitive to the empirical proxies used for periods when carbon-related risks become economically significant. To address this, the analysis evaluates robustness across alternative conditioning specifications using a broad systemic stress proxy (CISS) alongside transition-specific conditioning variables (EUA shocks and TRI), while consistently focusing on carbon exposure as the primary explanatory variable. This approach aligns with the literature emphasizing state-dependent and tail-risk manifestations of carbon risk (Ilhan et al. 2021; Bose et al. 2025).

5. Results

5.1. Sample, Descriptive Statistics, and Preliminary Diagnostics

The dataset comprises a balanced quarterly panel of 238 European firms observed from 2018 to 2024 (28 quarters), yielding 6664 firm-quarter observations, or 952 observations per year. Table 2 reports descriptive statistics for the primary variables.

Quarterly TotalReturn exhibits substantial dispersion (mean $\approx 2.8\%$, standard deviation $\approx 14.4\%$) and excess kurtosis (≈ 5.0), indicating fat-tailed behavior relative to a Gaussian benchmark. ESG pillar scores and the composite ESG score show moderate skewness and kurtosis (≈ 2 – 3), consistent with bounded yet persistent distributions. In contrast, CarbonScore is highly right-skewed and heavy-tailed (skewness > 8 , kurtosis > 86), reflecting concentration at low values and rare extreme outcomes. Because the sample is restricted to ESG Leaders firms with consistent emissions coverage, Appendix C.1 shows that raw CarbonScore is concentrated near zero (CarbonScore = 0 in 14.7% and CarbonScore ≤ 0.01 in 86.8% of observations), motivating within-year percentile ranking and Sector $\times$ Year standardization, and that the transformed CarbonScore_pct_st_SY remains well-behaved with limited tail mass.

Systemic stress indicators and market variables also deviate from Gaussianity; for example, CISS exhibits high skewness and kurtosis, whereas EURIBOR-3M shows pronounced left skewness and excess kurtosis. Annual summaries reveal that extreme returns and stress indicators cluster in specific subperiods: quarterly returns display substantially higher dispersion in 2020 (standard deviation $\approx 21.9\%$, min $\approx -63\%$, max $\approx +108\%$) and renewed volatility in 2022, whereas ESG scores remain relatively stable across years.

Similar time variation is observed in transition-related conditioning variables. EUA exhibits pronounced volatility across subperiods—particularly during 2020–2022—consistent with episodic repricing of carbon costs in the European emissions trading system, while TRI displays lower dispersion but persistent deviations from normality across years, reflecting sustained variation in the transition-related information environment. Because CISS, EUA, and TRI vary only over time, they are not interpreted as direct return determinants but are used exclusively as conditioning variables in interaction-based specifications. These distributional features motivate the use of outlier-robust preprocessing and dependence-robust inference in subsequent analyses.

Stationarity diagnostics indicate that financial returns and state variables behave as stationary processes at the quarterly frequency, whereas sustainability indicators, particularly ESG scores, exhibit high persistence consistent with near-unit-root behavior. LLC tests strongly reject the null of a unit root for all variables, including TotalReturn, ESG

scores, carbon measures, and macro-financial indicators, with adjusted statistics significant at the 1% level. IPS tests yield mixed results: for market and state variables (MSCI Europe returns, EURIBOR-3M, CISS, EUA, and TRI), the null of a unit root is decisively rejected; by contrast, for ESGScore and its E, S, and G pillars, the IPS statistics are positive with high p -values (≈ 0.88 – 0.97), failing to reject the null. For bounded or low-variation series, such as TotalReturn, CarbonScore, and controversy levels, the IPS test reports insufficient time periods to compute the W-t-bar statistic. Fisher-type ADF tests reinforce this heterogeneity, rejecting the null for returns and macro-financial variables while ESG scores and pillars consistently fail to reject it. Accordingly, we rely on fixed effects with dependence-robust inference and apply stationarity-friendly transformations for slow-moving sustainability indicators (within demeaning and detrending) rather than mechanical differencing.

Subsequent diagnostics confirm the appropriateness of the transformation steps implemented in Sections 4.1–4.3. The presence of heavy tails and excess kurtosis supports winsorization and rank-based transformations. Mixed stationarity, particularly for slow-moving ESG variables, justifies within-firm demeaning, firm-specific detrending, and percentile-based carbon measures instead of mechanical differencing. These procedures ensure that estimations rely on well-behaved within-firm variation while preserving the economic interpretability of sustainability characteristics.

5.2. Estimator Selection and Baseline Two-Way Fixed-Effects Results

Before presenting baseline estimates, fixed-effects and random-effects specifications of the panel model were compared. Quarter fixed effects (i.qdate) control for common time shocks. Three standard diagnostics are employed: (i) the fixed-effects test for joint significance of firm effects, (ii) the Breusch–Pagan Lagrange Multiplier test for random effects, and (iii) the Hausman test comparing fixed and random effects. The fixed-effects test yields $F(237, 6396) = 0.74$ ($p = 0.9987$) for the null that all firm effects are zero. The Breusch–Pagan LM test reports $\text{chibar2}(01) = 0.00$ ($p = 1.0000$), with a random-effects variance component of zero. The Hausman test reports $\chi^2(3) = 2.25$ ($p = 0.5224$), noting that $(V_b - V_B)$ is not positive definite.

Despite the Hausman test failing to reject the random-effects specification and the Breusch–Pagan LM test indicating a negligible variance component, the two-way fixed-effects specification (firm and quarter fixed effects) is retained as the baseline. This choice aligns with the empirical identification strategy in Section 4 and addresses potential correlations between time-invariant firm heterogeneity and sustainability characteristics, including carbon exposure, ESG scores, and controversy states. Reporting fixed-effects estimates ensures that all functional-form and state-dependence specifications rely on within-firm variation, with inference based on firm-clustered standard errors¹⁵.

Table 3, Column 1, presents the baseline two-way fixed-effects regression with firm and quarter fixed effects and firm-clustered standard errors. The model explains a substantial proportion of within-firm variation (within $R^2 = 0.3240$), and the joint F-test of included regressors and time effects is significant ($F(30, 237) = 45.97$, $p < 0.001$).

In this specification, the standardized carbon-exposure measure (CarbonScore_pct_st_SY) is positively associated with quarterly returns but not statistically significant. By contrast, the high-controversy indicator (ControvHigh) enters negatively and is statistically significant, whereas the standardized ESG composite (ESGScore_st_SY) is negative with marginal significance. The coefficient on ControvHigh corresponds to an estimated difference of roughly one percentage point in quarterly returns between high- and low-controversy firm-quarters, conditional on firm and time fixed effects. Variable definitions and preprocessing, including percentile ranking, Sector $\times$ Year standardization, and indicator construction, follow Section 4 and are summarized in the notes to Table 3.

Table 3. Main fixed-effects results (firm FE & quarter FE; firm-clustered SE).

Variable	M1	M2	M3	M4	M5
	Baseline	Carbon ²	Carbon Quintiles	Carbon × CISS	Quintiles CISS
CarbonScore_pct_st_SY	0.00124 (0.00328)	−0.00043 (0.00317)	—	0.00128 (0.00328)	—
CarbonScore_pct_st_SY ²	—	0.00559 (0.00255)	—	—	—
Q2	—	—	−0.01202 (0.00907)	—	−0.01145 (0.00904)
Q3	—	—	−0.02633 (0.01065)	—	−0.02554 (0.01061)
Q4	—	—	−0.01535 (0.01061)	—	−0.01509 (0.01059)
Q5	—	—	−0.00749 (0.01023)	—	−0.00702 (0.01022)
CISSz	—	—	—	0.03323 (0.01020)	0.02145 (0.01174)
Carbon × CISS	—	—	—	0.00310 (0.00233)	—
Q2 × CISS	—	—	—	—	0.02728 (0.00721)
Q3 × CISS	—	—	—	—	0.01101 (0.00667)
Q4 × CISS	—	—	—	—	0.00768 (0.00662)
Q5 × CISS	—	—	—	—	0.01451 (0.00711)
ControvHigh	−0.00964 (0.00427)	−0.00961 (0.00425)	(included)	−0.00975 (0.00428)	(included)
ESGScore_st_SY	−0.00735 (0.00405)	−0.00754 (0.00404)	(included)	−0.00729 (0.00406)	(included)
Observations (firm–quarter)	6664	6664	6664	6664	6664
Groups	238	238	238	238	238
Within R ²	0.3240	0.3247	0.3254	0.3245	0.3294

Source: Author estimates from panel regressions estimated in Stata/MP 17.0 using the cleaned quarterly firm-level panel described in Table 2. Note: Coefficients are reported with firm-clustered standard errors in parentheses. All specifications include firm and quarter fixed effects. CarbonScore_pct_st_SY denotes the Sector × Year standardized percentile carbon exposure measure; CarbonScore_pct_st_SY² is its squared term. Q2–Q5 represent carbon quintile indicators (Q1 omitted). A dash (—) indicates that the variable is not included. “Included” denotes absorbed controls. Observations (firm–quarter) denote firm–quarter observations.

To assess economic magnitude, we translate key coefficients into quarterly percentage-point effects. In the baseline linear specification (Table 3, M1), a one-standard-deviation increase in the standardized carbon-exposure measure is associated with approximately +0.12 percentage points in quarterly returns (0.00124 in return units) and is statistically indistinguishable from zero. By contrast, entering a high-controversy state is associated with roughly −0.96 percentage points lower quarterly returns (−0.00964), while a one-standard-deviation increase in the standardized ESG composite is associated with roughly −0.74 percentage points lower quarterly returns (−0.00735). In the quintile–stress specification (Table 3, M5), a one-standard-deviation increase in systemic stress (CISS) is associated with approximately +2.1 percentage points in quarterly returns for the low-carbon reference group, with larger stress sensitivities estimated for some carbon quintiles. Functional-form and stress-conditioning results are reported in Sections 5.3 and 5.4.

5.3. Carbon Functional-Form Specifications

Table 3 examines whether the carbon–return relationship depends on the functional form used to represent carbon risk, while retaining the two-way fixed-effects structure and firm-clustered inference. Across Columns 2 and 3, the overall model fit remains similar to the baseline, indicating that functional-form modifications do not materially affect explanatory power once firm and time effects are included.

Column 2 introduces a quadratic carbon term. The quadratic specification includes the standardized carbon exposure measure (CarbonScore_pct_st_SY) and its squared term (CarbonScore_pct_st_SY²) to allow for nonlinear functional-form effects. The quadratic component is statistically significant, whereas the linear term remains indistinguishable from zero. This pattern indicates nonlinearity in the carbon–return relationship: the association varies across the carbon distribution, with the marginal effect changing as exposure increases. The data are more consistent with curvature than with a constant linear slope.

Column 3 replaces the continuous carbon measure with carbon quintile indicators, allowing for a flexible nonparametric comparison of returns across carbon-exposure groups.

Estimates do not display a monotonic ranking across quintiles; instead, differences are concentrated in the middle of the distribution, with the third quintile (Q3) differing from the low-carbon reference group while the remaining quintiles are not individually distinguishable. A joint test of the quintile block is marginal ($p \approx 0.057$), suggesting that the discrete specification is marginal in the baseline definition to the baseline once fixed effects are included. Collectively, Columns 2–3 indicate that carbon-related return differences are not well captured by a simple linear model and are better described by localized variations across the carbon distribution rather than a smooth monotonic premium. As an additional robustness check against industry composition and within-universe range restriction, we reconstruct carbon quintiles within Sector $\times$ Year cells and re-estimate the fixed-effects model. Under this alternative binning, the quintile block becomes jointly significant ($F(4, 237) = 2.73$; $p = 0.030$; Appendix C.4). We further absorb Sector $\times$ Quarter fixed effects (firm FE + Sector $\times$ Quarter FE) to control for time-varying industry shocks; the linear carbon coefficient remains economically small and statistically indistinguishable from zero ($\beta = 0.00119$; $p = 0.716$; Appendix C.5).

5.4. Systemic Stress Interaction Specifications

Table 3 (Columns 4–5) examines whether the return association of carbon exposure is state-dependent, using the standardized systemic stress indicator CISSz. Both specifications retain the two-way fixed-effects structure and firm-clustered inference of the baseline, so differences across columns reflect variations in how stress enters the model and interacts with carbon exposure.

Column 4 introduces systemic stress in a continuous interaction framework by including CISSz and the Carbon $\times$ Stress interaction term. The results indicate that systemic stress is positively related to returns, while the Carbon $\times$ Stress interaction is not statistically different from zero. This suggests that, under a linear interaction assumption, the carbon–return relationship does not systematically change with stress conditions. The main effect of carbon remains small, and including stress does not materially alter overall model fit relative to earlier specifications.

Column 5 relaxes the linear interaction assumption by allowing stress sensitivity to vary across carbon quintiles. Compared with Column 4, this specification provides direct evidence of heterogeneity: the interaction terms differ across quintiles and are jointly significant ($F(4, 237) = 4.14$, $p = 0.0029$), implying that return sensitivity to systemic stress is not uniform across the carbon distribution. Marginal effects¹⁶ indicate that the estimated impact of systemic stress on returns is positive across all quintiles, strongest in the second quintile, and generally statistically significant for most quintiles, with the lowest-carbon group estimated least precisely. This quintile-based interaction suggests that stress exposure is not adequately captured by a single Carbon $\times$ Stress slope but varies across carbon categories.

To determine whether the limited evidence for conditional carbon pricing under systemic financial stress is influenced by the choice of conditioning proxy, parallel interaction specifications are estimated using transition-specific salience measures. Appendix B.1 presents continuous interaction models in which carbon exposure is interacted with EU ETS allowance price shocks and a news-based Transition Risk Index, in addition to the CISS specification. Appendix B.2 extends this analysis by employing carbon–quintile interactions to capture nonparametric heterogeneity. In the continuous specification, the Carbon $\times$ EUA interaction is positive and marginally significant, whereas the Carbon $\times$ TRI interaction is statistically insignificant. In the quintile-based specifications, the EUA interaction block is jointly significant ($p \approx 0.024$), whereas the TRI interaction block is not ($p \approx 0.253$), indicat-

ing heterogeneous transition-salience responses across carbon groups without a stable or monotonic carbon penalty.

5.5. Additional Inference Specifications and Diagnostic Outputs

Table 4 summarizes robustness checks and alternative inference approaches for the baseline specification (Table 3, Column 1) and the quintile–stress specification (Table 3, Column 5). The purpose is to evaluate whether the main conclusions depend on assumptions about the error structure, the timing of sustainability characteristics, or alternative methods for modeling unobserved firm heterogeneity. In addition to the CISS-based stress specifications presented in Table 4¹⁷, robustness to alternative transition-risk conditioning proxies is examined using EU ETS allowance price shocks and a news-based Transition Risk Index. Table 4 summarizes the key interaction coefficients and joint tests for these transition-conditioned specifications, while the corresponding continuous and quintile-based models are reported in full in Appendix B. Taken together, these results confirm that the principal findings are robust to the choice of conditioning variable: transition-specific salience does not yield a stable or monotonic negative carbon premium in mean returns. To control for sector-specific quarter shocks, we re-estimate the baseline model replacing quarter fixed effects with Sector $\times$ Quarter fixed effects (firm FE and Sector $\times$ Quarter FE). The linear carbon coefficient remains economically small and statistically indistinguishable from zero ($\beta \approx 0.0012$; $p \approx 0.72$), indicating that the baseline null is not explained by time-varying industry shocks. Results are reported in Appendix C.5.

Dependence-robust inference. The first robustness check replaces firm-clustered standard errors with a dependence-robust variance estimator that accommodates both serial correlation and cross-sectional dependence. Table 4 shows that this adjustment does not materially change the patterns observed in Table 3. In the baseline model, the carbon coefficient remains economically small and statistically insignificant, while the controversy indicator remains negative and significant. The ESG composite is also negative and estimated with greater precision. In the quintile–stress model, heterogeneity in stress sensitivity is concentrated in specific interaction terms, indicating that the observed interaction pattern is not driven by the choice of clustering. Table 4 provides summary evidence for transition-risk conditioning via EUA shocks and TRI, while the full interaction specifications are detailed in Appendix B. The marginal significance of the continuous Carbon $\times$ EUA interaction does not contradict the significant quintile $\times$ EUA block; the latter reflects heterogeneous EUA sensitivity across carbon groups rather than a single linear conditional premium.

Timing sensitivity via lagged characteristics. Table 4 also reports results incorporating one-quarter lags of primary firm-level characteristics, reducing the number of usable observations. The lagged results preserve the qualitative patterns of the baseline: the controversy indicator remains negative and significant, the ESG composite remains negative and significant, and the carbon coefficient remains economically small and statistically indistinguishable from zero. These results suggest that the baseline findings do not depend on contemporaneous measurement of sustainability characteristics.

Alternative clustering level. To assess sensitivity to coarser dependence structures, Table 4 reports the quintile–stress specification with standard errors clustered at the country level. Compared with firm-level clustering, this adjustment increases uncertainty and reduces statistical significance for some interaction terms. The strongest interaction terms remain significant, whereas others lose precision under the coarser clustering. This indicates that inference regarding heterogeneous stress effects is more sensitive to clustering level than are the baseline controversy and ESG effects.

Table 4. Robustness and alternative inference.

Robustness/Check	Alternative Inference/Specification	Key Coefficient(s) Reported	Estimate (SE), <i>p</i> -Value
Dependence-robust SE	Driscoll–Kraay FE (lag 4), baseline	CarbonScore_pct_st_SY	0.00124 (0.00338), <i>p</i> = 0.718
		ControvHigh	−0.00964 (0.00232), <i>p</i> < 0.001
		ESGScore_st_SY	−0.00735 (0.00290), <i>p</i> = 0.017
Dependence-robust SE	Driscoll–Kraay FE (lag 4), quintiles × CISS	Q2 × CISSz	0.02728 (0.00252), <i>p</i> < 0.001
		Q3 × CISSz	0.01101 (0.00458), <i>p</i> = 0.023
		Q4 × CISSz	0.00768 (0.00584), <i>p</i> = 0.199
		Q5 × CISSz	0.01451 (0.00359), <i>p</i> < 0.001
Lag structure	FE with one-quarter lags (cluster firm)	L1 CarbonScore_pct_st_SY	0.00244 (0.00325), <i>p</i> = 0.453
		L1 ControvHigh	−0.00973 (0.00418), <i>p</i> = 0.021
		L1 ESGScore_st_SY	−0.00945 (0.00400), <i>p</i> = 0.019
Alternative clustering	FE quintiles × CISS (cluster country; 16 clusters)	Q2 × CISSz	0.02728 (0.00764), <i>p</i> = 0.003
		Q5 × CISSz	0.01451 (0.01055), <i>p</i> = 0.189
Correlated RE (optional)	Random effects augmented with firm means (Mundlak)	Carbon_mean	−0.00112 (0.00343), <i>p</i> = 0.744
		ESG_mean	−0.00296 (0.00155), <i>p</i> = 0.056
		ControvHigh	−0.00654 (0.00275), <i>p</i> = 0.018
Two-way clustering attempt (optional)	Two-way clustered SE (firm & quarter)	SE availability	Variance estimator singular/ill-conditioned; SE not available
Transition-risk proxies: continuous interaction models (Appendix B.1)	Two-way FE interaction models (firm FE + quarter FE; firm-clustered SE)	Carbon × EUA	0.00262 (0.00154), <i>p</i> ≈ 0.090
		Carbon × TRI	−0.00184 (0.00169), <i>p</i> ≈ 0.276
Transition-risk proxies: quintile interaction blocks (Appendix B.2)	Two-way FE quintile interaction models (firm FE + quarter FE; firm-clustered SE)	Joint test: quintiles × EUA	F-test <i>p</i> ≈ 0.024
		Joint test: quintiles × TRI	F-test <i>p</i> ≈ 0.253

Source: Author estimates computed in Stata/MP 17.0 using the quarterly panel described in Table 2. Note: Reported estimates correspond to the specifications described in Section 5.5. Standard errors are firm-clustered unless otherwise stated. Driscoll–Kraay standard errors use a lag length of four quarters. *p*-values refer to two-sided tests. “SE availability” indicates cases where the variance estimator was singular or ill-conditioned and standard errors could not be computed. L1 denotes one-quarter lagged variables. Carbon_mean and ESG_mean denote firm-level averages used in the correlated random-effects (Mundlak) specification. Joint tests refer to F-tests of the indicated interaction blocks. Carbon × EUA and Carbon × TRI are coefficients from continuous interaction models; the joint tests refer to separate quintile interaction models and capture heterogeneity across carbon groups rather than a single linear interaction slope.

Correlated random effects (Mundlak) and two-way clustering feasibility. Table 4 also presents a correlated random-effects (Mundlak) specification, which separates within-firm and between-firm components by including firm means of key regressors. The firm-mean terms suggest that the between-firm component is more relevant for ESG than for carbon, while the controversy indicator remains negative and significant. Attempts to compute two-way clustered standard errors (by firm and quarter) did not produce usable estimates due to numerical instability; therefore, two-way clustering is not adopted as a primary inference benchmark. Baseline inference uses firm-clustered standard errors, with Driscoll–Kraay standard errors as the dependence-robust alternative.

Residual diagnostics. Table 5 reports diagnostics for serial correlation, cross-sectional dependence, heteroskedasticity, and collinearity. The results indicate that serial correlation and groupwise heteroskedasticity are significant, supporting the use of cluster-robust and dependence-robust inference in Table 4. Evidence on cross-sectional dependence is mixed across tests. Collinearity diagnostics, based on variance inflation factors, yield mean VIF values of 1.84 for baseline regressors and 5.53 for interaction regressors.

Table 5. Residual diagnostics for serial correlation, cross-sectional dependence, heteroskedasticity, and collinearity.

Diagnostic	Statistic	p-Value/Reference
Wooldridge test for autocorrelation (baseline regressor set)	$F(1, 237) = 30.078$	$p < 0.001$
Wooldridge test for autocorrelation (regressor set incl. CISSz)	$F(1, 237) = 1.314$	$p = 0.2529$
Pesaran test for cross-sectional dependence	CD = -1.860 Avg. absolute off-diagonal correlation = 0.204	$p = 0.0628$ —
Friedman test for cross-sectional dependence	13.075	$p = 1.0000$
Frees test for cross-sectional dependence	4.563	Crit. values: 0.0924 (10%), 0.1204 (5%), 0.1726 (1%)
Modified Wald test for groupwise heteroskedasticity	$\chi^2(238) = 3132.98$	$p < 0.001$
Variance Inflation Factors (OLS with quarter indicators)	Mean VIF = 1.84 (baseline); 5.53 (interaction models)	—

Source: Author estimates computed in Stata/MP 17.0 using the quarterly panel described in Table 2. Note: Reported statistics correspond to diagnostic tests applied to the baseline fixed-effects specification unless otherwise indicated. Wooldridge tests evaluate first-order autocorrelation in panel data. Pesaran, Friedman, and Frees tests assess cross-sectional dependence. The Modified Wald test evaluates groupwise heteroskedasticity across firms. Variance Inflation Factors (VIF) are computed from an OLS regression including quarter indicators; reported values denote the mean VIF for baseline and interaction specifications. Critical values for the Frees test are shown in parentheses. p-values refer to two-sided tests. Interaction models include carbon–state and quintile–state interaction specifications using CISS, EUA shocks, and TRI.

6. Discussion

This section evaluates the empirical evidence presented in Section 5 in relation to the literature reviewed in Section 2 and the main hypothesis, H1. The two-way fixed-effects model isolates within-firm changes in carbon exposure and sustainability characteristics while controlling for unobserved, time-invariant firm heterogeneity and common quarter shocks. Three principal findings emerge. First, carbon exposure does not exhibit a consistent linear association with returns in either the baseline specification or robustness checks. Second, when the model incorporates nonlinear functional forms and heterogeneous stress sensitivities, carbon-related effects become apparent but are non-monotonic and do not conform to a simple “higher carbon → lower return” relationship. Third, although ESG and controversy variables are included as controls, both display economically significant and

statistically robust patterns consistent with measurement-noise and event-risk mechanisms documented in prior research.

In addition to documenting the absence of a stable linear carbon premium, this study clarifies the conditions and mechanisms through which carbon-related information is incorporated into equity returns within an investable European ESG universe, employing conservative within-firm identification. The use of a strongly balanced panel, firm and quarter fixed effects, and the separation of carbon exposure from ESG composites and controversy states enables the distinction between (i) slow-moving structural transition exposure, (ii) preference- and measurement-driven ESG characteristics, and (iii) event-driven downside shocks. Functional-form flexibility, through quadratic and quintile designs, addresses threshold and segmentation mechanisms that a linear premium would obscure. Stress interactions are used to operationalize conditional pricing, aligning with regime-dependent risk-bearing capacity and attention-based repricing. Collectively, these methodological choices refine the interpretation from “carbon is not priced” to “carbon pricing is nonlinear and state-dependent”, whereas controversy-related downside risk is the most consistently return-relevant sustainability signal in this sample.

6.1. Nonlinear Carbon Effects and Reconciling Mixed Evidence

The functional-form results indicate that carbon exposure is not adequately summarized by a single linear premium. In the quadratic specification (Table 3, Column 2), the linear carbon coefficient remains statistically indistinguishable from zero, while the squared term is significant ($\text{CarbonScore}_{pct,st_SY}^2 = 0.00559$, $p = 0.030$). This finding suggests curvature in the within-firm carbon–return relationship: return differences associated with carbon exposure vary across the carbon distribution rather than remaining constant per unit of carbon. Economically, this curvature is consistent with threshold-like mechanisms, where carbon exposure becomes return-relevant primarily for firms at higher exposure levels, for instance, due to regulatory attention, financing constraints, or screening cut-offs. The financing-constraint channel aligns with empirical evidence from credit markets. Dong et al. (2025) identify a carbon-risk premium in U.S. syndicated loan pricing, demonstrating that higher carbon intensity is associated with increased loan risk spreads. The magnitude of this premium depends on the environmental commitments of both borrowers and lenders and intensifies during periods of monetary tightening. This state-dependent repricing of debt financing costs supports the existence of a discount-rate, or cost-of-capital, channel through which carbon exposure becomes increasingly relevant to returns, especially when financing conditions become more restrictive. This pattern aligns with mechanisms where transition risk is priced only after exposures surpass salience or constraint thresholds, such as regulatory scrutiny, financing constraints, investor screening cutoffs, or convex adjustment costs. Consequently, a linear specification averages across regions with minimal incremental pricing.

The nonparametric quintile specification reinforces this interpretation. Carbon quintiles do not produce a monotone ranking in returns (Table 3, Column 3). Relative to the low-carbon reference group (Q1), the largest and statistically significant difference occurs in the middle of the distribution (Q3 = -0.02633 , $p = 0.014$), whereas Q2, Q4, and Q5 are not individually distinguishable from Q1, and the joint quintile test is marginally significant ($p \approx 0.057$). This result indicates that carbon sorting does not produce a smooth brown-minus-green return spread in this sample. Non-monotonicity offers a plausible explanation for the mixed signs and unstable magnitudes reported in the carbon-premium literature: if pricing is localized or segmented by investor clienteles, linear specifications average across offsetting regions and yield weak or inconsistent slopes.

The observed non-monotonic pattern—where return differentials are concentrated in the middle of the carbon distribution—does not imply that intermediate emitters are inherently more (or less) profitable than high- or low-carbon firms. Instead, the evidence is consistent with localized discount-rate repricing, where the market's required return responds most strongly for firms positioned near economically salient thresholds within a screened investable universe. This interpretation aligns with theories of sustainable investing in which investor tastes, ESG mandates, and screening constraints shape demand and equilibrium pricing (e.g., [Pástor et al. 2021](#); [Pedersen et al. 2021](#)).

- (i) **Investor screening and “marginal” reclassification.** In an ESG-screened universe, extreme high-carbon firms are partly truncated and often already treated as transition-risk exposures, while low-carbon firms are perceived as transition-resilient. Firms in the mid-range are more likely to sit near screening cutoffs or mandate constraints; relatively small changes in emissions and/or disclosure can shift investor eligibility and generate stronger price pressure than at the extremes ([Pedersen et al. 2021](#); [Ehlers et al. 2021](#)). Consistent with the disclosure-based component of this channel, evidence from China shows that environmental information disclosure can increase stock price informativeness by embedding more information about future cash flows in prices ([Yang et al. 2025](#)).
- (ii) **Regulatory salience and convex or threshold-based compliance costs.** Transition policies can introduce nonlinearities because costs may rise disproportionately when firms approach points of elevated scrutiny or binding constraints (e.g., cap-and-trade designs that require purchasing allowances beyond an emissions limit) ([Ehlers et al. 2021](#)).
- (iii) **Heterogeneous technological adjustment and uncertainty.** Intermediate emitters can face greater uncertainty regarding the feasibility, timing, and credibility of abatement pathways, making discount rates more sensitive to incremental carbon information. This channel is consistent with evidence that carbon-transition risk premia are tied to uncertainty about technological progress and policy support and can vary over time with shifts in beliefs and salience ([Bolton and Kacperczyk 2023](#)).

Importantly, these mechanisms imply that repricing pressure is strongest for “marginal” emitters whose investor eligibility, regulatory exposure, or perceived transition feasibility can change discontinuously in response to relatively small updates in emissions or disclosure. Taken together, these channels imply segmented and locally concentrated pricing, which helps explain why linear specifications can produce weak average premia even when curvature and mid-range differences are statistically detectable.

These results align with cross-sectional evidence documenting carbon-related return differences in broader universes ([Bolton and Kacperczyk 2021, 2023](#)). The present setting differs in two ways that reduce the likelihood of detecting a monotone gradient: (i) firm fixed effects absorb persistent cross-firm differences in business models and baseline emissions intensity and (ii) the ESG Leaders universe compresses dispersion among high-emissions firms compared to a full-market sample. Under these conditions, carbon-related pricing is more likely to emerge through nonlinearities and conditional sensitivities than through a constant linear premium.

6.2. Carbon Exposure and Average Return Premium

The baseline two-way fixed-effects estimates provide clear evidence supporting the first component of H1. Carbon exposure is economically negligible and statistically insignificant in the baseline model (Table 3, Column 1: 0.00124, $p = 0.706$), remains insignificant under dependence-robust Driscoll–Kraay inference (Table 4: $p = 0.718$), and is still insignificant when lagged by one quarter (Table 4: $p = 0.453$). Within this quarterly,

within-firm framework, carbon exposure does not behave as a stable priced characteristic in mean returns. The within-firm fixed-effects design is intentionally conservative. It eliminates persistent differences between “brown” and “green” firm types, focusing on repricing due to changes in exposure rather than static business-model differences. This approach helps explain the null mean-return result considering broader cross-sectional carbon-premium evidence.

This result is consistent with studies suggesting that carbon risk manifests less through average return differentials and more through tail, crash-risk, and financing channels. Ilhan et al. (2021) show that carbon risk is reflected in tail-risk pricing, and Bose et al. (2025) link carbon risk to future crash risk. Other studies document risk and valuation channels beyond unconditional equity-return premia (Al Rabab’a et al. 2024; Duppati 2025), while evidence from credit markets indicates that carbon intensity can affect financing conditions even when return premia are weak or unstable (Ehlers et al. 2021; Dong et al. 2025). Thus, the baseline null is economically interpretable rather than puzzling: carbon risk appears to be priced primarily through nonlinear or downside mechanisms not fully captured by contemporaneous quarterly mean-return regressions.

6.3. Conditional Carbon Pricing and Systemic Stress

The second component of H1 predicts that carbon exposure is negatively priced during periods when carbon-related risks are economically significant. Empirically, such periods are proxied by macro-financial systemic stress ($CISS_z$). The results provide limited evidence for a negative conditional carbon penalty under this proxy.

Under the continuous interaction specification (Table 3, Column 4), the Carbon $\times$ CISS interaction is not statistically significant (0.00310, $p = 0.186$), indicating no systematic linear change in the carbon–return relationship as systemic stress rises. Relaxing this restriction by interacting stress with carbon quintiles (Table 3, Column 5) yields a jointly significant interaction block ($F(4, 237) = 4.14$, $p = 0.0029$), demonstrating that return sensitivity to stress varies across carbon groups. However, the most precisely estimated interactions (notably Q2 $\times$ CISS and Q5 $\times$ CISS) are positive, implying a positive stress–return slope across quintiles, strongest in Q2. This pattern does not support a uniform “higher carbon becomes more negatively priced during stress” interpretation.

Two implications follow. First, state dependence exists but is better characterized as heterogeneous stress sensitivity across carbon categories rather than as a uniform negative carbon penalty. Second, CISS, while a standard measure of systemic financial stress, is not a direct measure of transition-risk conditions. Systemic stress serves as a proxy for risk-bearing capacity rather than representing a transition-specific shock. A natural reason is that CISS captures broad system-wide financial stress and risk-bearing capacity, not transition-specific repricing triggers. Transition salience can instead be proxied by market-based carbon-cost shocks (e.g., EU ETS allowance price innovations) or by policy-event windows tied to major EU climate-policy milestones (e.g., the European Green Deal and Fit for 55 legislative-package announcements and subsequent EU ETS reform steps). Conditioning on such transition-specific salience measures can better isolate whether and when carbon exposure becomes return-relevant, relative to conditioning on broad systemic stress.

When carbon repricing is primarily driven by transition-policy news, carbon-price shocks, or increased climate attention, conditioning on broad systemic stress may produce heterogeneous responses. This does not necessarily indicate a uniform carbon penalty under stress. Systemic stress can coincide with heterogeneous sectoral and macro-financial shocks whose relationship to carbon exposure is not uniformly negative. Therefore, conditional carbon pricing may require variables more closely tied to transition risk, such as carbon-price shocks, climate-policy announcements, or transition-risk news indices. This interpretation

aligns with time-varying approaches emphasizing that ESG-related pricing effects depend on attention and regime dynamics (Alessi et al. 2023) and with equilibrium models in which investor preferences shift expected returns without implying a constant negative premium (Pástor et al. 2021). Overall, the carbon results robustly support H1's "no average association" component but provide weak support for the claim that carbon becomes negatively priced during periods of heightened systemic stress. Instead, carbon-related effects are more visible through nonlinear exposure patterns and heterogeneous state sensitivities, consistent with the literature emphasizing conditional and tail-risk manifestations of carbon risk. Importantly, this conclusion is not driven by an overly coarse conditioning proxy: even when conditioning directly on transition-specific salience—using EU ETS allowance price shocks and a news-based Transition Risk Index—the return sensitivity of carbon exposure remains heterogeneous and non-monotonic. Specifically, carbon-price shocks elicit at most weak, proxy-dependent responses, whereas transition-risk news does not systematically amplify carbon-related return penalties. This evidence suggests that transition risk is not incorporated into equity returns through a simple, contemporaneous discount-rate channel tied to observable policy or price signals. Instead, carbon-related repricing appears episodic, localized, and mediated by institutional constraints, investor segmentation, or financing channels, consistent with recent work emphasizing attention, tail risk, and nonlinear adjustment mechanisms rather than stable transition premia.

6.4. ESG Composite as a Control: Evidence of a "Lower Required Return/Greenium" Pattern

Although the ESG composite score is included primarily as a control for broad sustainability and reputational signals, its coefficient is informative. $ESGScore_{st_SY}$ is consistently negative across specifications and becomes statistically significant under Driscoll–Kraay inference and in lagged regressions (Table 4: $p = 0.017$; one-quarter lag $p = 0.019$). This pattern aligns with models of sustainable investing, in which demand from sustainability-oriented or constrained investors bids up prices of high-ESG firms, lowering expected returns (Pástor et al. 2021). The coefficient's marginal significance under baseline firm-clustered inference and improved precision under Driscoll–Kraay inference is consistent with the literature on measurement noise and rating disagreement (Berg et al. 2022; Christensen et al. 2022; Avramov et al. 2022). According to this interpretation, the negative association between ESG and returns reflects a demand or valuation channel rather than compensation for systematic risk. Therefore, ESG composites are treated as controls instead of serving as the primary proxy for transition risk.

A second interpretation, consistent with the negative sign, is that improvements in composite ESG reduce perceived downside risk, lowering required returns. In either case, ESG captures a nontrivial pricing-relevant component in this sample, though it does not support a stable positive "ESG alpha" narrative.

6.5. Controversy Risk as a Control: The Most Robust Return-Relevant Sustainability Signal

The controversy indicator produces the most robust and economically interpretable sustainability result. High-controversy firm-quarters underperform by approximately one percentage point per quarter in the baseline specification (Table 3, Column 1: $ControvHigh = -0.00964$, $p = 0.025$), with stronger significance under Driscoll–Kraay inference (Table 4: $p < 0.001$) and persistence in the one-quarter lag specification (Table 4: -0.00973 , $p = 0.021$). This stability is consistent with controversies reflecting event-driven downside exposure, distinct from slow-moving ESG characteristics (Serafeim and Yoon 2022; Bang et al. 2023). Controversies likely capture abrupt negative information, such as regulatory violations, litigation, environmental incidents, or governance failures, which can trigger asymmetric losses and mandate-constrained investor rebalancing. The persistence

of controversy effects aligns with an information-shock mechanism, in which controversies condense sustainability information into discrete events. This process generates short-term price pressure and asymmetric downside risk that are not fully captured by gradually updated emissions data and composite ESG scores.

This result is crucial for interpreting carbon and ESG coefficients: controlling for controversy reduces the risk that episodic negative ESG events are misattributed to structural carbon exposure or composite ESG characteristics. In contexts where sustainability information arrives partly through discrete “bad news” events, controversy controls are central to separating slow-moving exposure measures from shock-like sustainability realizations.

Overall, the evidence suggests that carbon-related pricing among European ESG leaders does not follow a stable linear premium. Rather, the association between returns and carbon exposure is concentrated in specific segments of the exposure distribution and varies according to state variables. This pattern aligns with threshold-based and regime-dependent repricing, as opposed to a continuous brown-minus-green gradient. From a methodological perspective, these findings highlight the necessity of distinguishing structural exposure, such as emissions, from broader ESG composites and event-driven controversy states, while employing flexible functional forms under dependence-robust inference. Substantively, the results indicate that investors seeking carbon premia within this ESG universe should anticipate conditional and non-monotonic return patterns. Controversy risk, however, remains the most consistently return-relevant sustainability signal.

7. Conclusions

This study examines whether firm-level carbon exposure is conditionally priced in European equity returns using a balanced quarterly panel of 238 firms from the MSCI Europe ESG Leaders universe over the period 2018–2024. The empirical framework employs two-way fixed-effects regressions at the firm and quarter levels with dependence-robust inference to test H1. The analysis distinguishes structural carbon exposure from broader ESG characteristics and event-driven controversy states.

The central finding is that, within firms, carbon exposure does not operate as a stable, linearly priced return characteristic in quarterly mean returns. Across the baseline specification and key robustness checks, including dependence-robust standard errors and one-quarter lags, the estimated linear carbon coefficient remains economically small and statistically indistinguishable from zero. A consistent “carbon premium” therefore does not emerge in this within-firm, quarterly setting when carbon exposure enters linearly. This result supports the first part of H1 and is consistent with prior evidence that climate-transition exposure may be reflected less in average returns and more through state dependence or downside-risk channels (Bolton and Kacperczyk 2021, 2023; Ilhan et al. 2021).

The absence of a linear premium does not imply a lack of structure in carbon-related return differences. Functional-form tests indicate pronounced nonlinearity. The quadratic carbon term is statistically significant even when the linear term is not, and a carbon-quintile specification yields a non-monotonic pattern, with return differentials concentrated in the middle of the carbon distribution rather than following a smooth brown-minus-green gradient. These results suggest that carbon-related pricing effects, if present, are localized or threshold-based rather than proportional to marginal changes in exposure. This finding helps reconcile the mixed evidence in the literature when analyses rely solely on linear specifications. Evidence for the second part of H1, which predicts negative pricing of carbon exposure when risks become economically salient, is weaker when salience is proxied by systemic financial stress. Using the ECB CISS, the interaction between continuous carbon exposure and stress is not statistically significant. Allowing stress sensitivity to vary across carbon quintiles produces a jointly significant interaction block,

indicating heterogeneity in stress responses across carbon groups. However, this pattern does not imply a uniform negative carbon penalty during high-stress periods. Overall, the results point to state dependence and heterogeneity, but systemic financial stress does not appear to capture transition-risk repricing directly. This conclusion aligns with research emphasizing time-varying ESG-related pricing and the importance of the conditioning information set (Alessi et al. 2023).

Although the composite ESG score and the controversy indicator primarily serve as controls, they yield additional insights. Controversy states are consistently associated with lower returns, corresponding to approximately one percentage point of quarterly underperformance in the baseline specification. This effect remains robust under dependence-robust inference and lagged models, consistent with evidence that markets respond strongly to material, event-driven ESG news (Serafeim and Yoon 2022; Bang et al. 2023). The ESG composite score generally exhibits a negative association with returns, which becomes more precisely estimated under stronger inference and lagged specifications. This pattern accords with preference-based models of sustainable investing, in which demand for greener firms reduces expected returns, and with evidence that composite ESG ratings are noisy and provider-dependent (Pástor et al. 2021; Berg et al. 2022).

The results indicate that carbon-related pricing within the investable European ESG Leaders universe does not exhibit a consistent linear premium in quarterly mean returns. Rather, the analysis reveals curvature, non-monotonicity, and varying state sensitivity across the carbon distribution. These findings suggest that transition policy, financial supervision, and investment implementation should not depend on linear “carbon beta” assumptions but instead adopt threshold- and regime-aware methodologies.

- (1) **Carbon-pricing design:** Prioritize predictability and salience rather than assuming a linear cost-of-capital channel. Given that equity-market return responses to carbon exposure are nonlinear and episodic, reliance on a uniform, contemporaneous “carbon premium” to influence emitters is likely ineffective, especially in screened, benchmark-driven universes. Transition policy can address nonlinear repricing and strengthen incentives by offering credible forward guidance on carbon cost trajectories, including transparent multi-year cap trajectories and rule-based adjustments. This strategy reduces uncertainty about when firms surpass economically significant thresholds.
- (2) **Disclosure and assurance:** Minimize threshold uncertainty that drives nonlinear pricing. Non-monotonic return patterns indicate that markets react more intensely near salient cutoffs, such as screening constraints, regulatory scrutiny, or credibility thresholds. This pattern highlights the necessity of consistent, assured emissions reporting and well-defined transition-plan milestones, including interim targets, capital expenditure alignment, and verified progress. Improved disclosure can decrease discontinuities in investor perceptions and help avert abrupt repricing near perceived thresholds.
- (3) **Financial supervision and stress testing:** Employ nonlinear models for carbon risk and integrate “event-risk” signals. Because stress sensitivity differs across carbon quintiles rather than following a single linear relationship, supervisory monitoring and climate stress tests should reflect nonlinear exposures, such as bucketed carbon states, rather than presuming a constant carbon beta. Additionally, the consistent negative relationship between high-controversy states and returns suggests that controversy indicators can serve as operational “early-warning” signals for downside sustainability shocks and should be incorporated into risk dashboards and scenario planning.
- (4) **Investors and index providers:** Apply nonlinear carbon-risk controls and explicitly consider controversy states in pricing. In screened universes, portfolio construction and risk models should not assume monotonic brown-minus-green spreads. A practi-

cal method is to use nonparametric carbon buckets, such as quintiles or thresholds, and to incorporate regime interactions in expected return and risk models. Controversy risk should be recognized as a separate, short-term factor that may have a greater impact than mean-return sustainability effects.

Several limitations qualify the interpretation of these findings. Applicability boundary: the conclusions are best interpreted as evidence for large- and mid-cap European firms within an ESG-screened, benchmark-driven universe with consistent emissions/ESG coverage, and they should not be treated as population-wide estimates for the full European equity market—particularly for high-emitting firms excluded by ESG screens or firms without reliable disclosure histories. The ESG Leaders screen likely compresses variation in carbon exposure by excluding much of the high-emissions segment of the market, while the within-firm fixed-effects design deliberately downweights between-firm channels central to cross-sectional carbon-premium studies. This range restriction can mechanically reduce the power to detect linear pricing effects and can shift any detectable structure toward nonlinear or threshold-like patterns. Consistent with this concern, Appendix C reports dispersion and within-between decompositions and shows that trimming extreme carbon observations or firms does not alter the baseline null linear carbon coefficient (Tables A3–A5). Carbon exposure is proxied by reported total emissions interpolated from annual disclosures to quarterly frequency, and the conditioning variable captures macro-financial stress rather than transition-specific shocks. In addition, the analysis focuses on quarterly mean returns. If carbon risk primarily affects tail risk, crash risk, or financing conditions, mean-return regressions may understate its economic relevance. Future research should therefore complement within-firm identification with cross-sectional designs and broader samples beyond ESG Leaders. It should condition on transition-specific salience measures, such as carbon-price shocks, climate-policy announcement windows, and transition-risk news indices, rather than broad systemic stress. Extending the outcome space beyond mean returns to encompass downside risk, tail risk, and financing channels would further clarify when carbon-related risks become economically significant and how they are transmitted into asset prices.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study were obtained from third-party licensed sources, including S&P Capital IQ Pro (<https://www.spglobal.com/marketintelligence/en/solutions/capital-iq-platform>) (accessed on 30 August 2025), Sustainalytics (<https://www.sustainalytics.com>, accessed on 30 August 2025), and DitchCarbon (<https://ditchcarbon.com>, accessed on 29 April 2025 and 20 December 2026), EU ETS allowance prices (data vendor/exchange) <https://tradingeconomics.com/commodity/carbon> (accessed on 29 January 2026), and from public institutional sources (ECB Statistical Data Warehouse) (<https://data.ecb.europa.eu/data/datasets/CISS/data-information?dataset%5B0%5D=Composite%20Indicator%20of%20Systemic%20Stress%20%28CISS%29&advFilterDataset%5B0%5D=Composite%20Indicator%20of%20Systemic%20Stress%20%28CISS%29&showDatasetModal=false>, accessed on 3 October 2025), and Transition Risk Index (TRI) dataset. Data are also available at <http://www.policyuncertainty.com> (accessed on 29 January 2026). Restrictions apply to the availability of these data due to provider licensing terms and the re-identifiable nature of firm-level financial and emissions data. The data are available from the respective data providers with their permission. Replication materials, including variable definitions, data construction procedures, and estimation codes, are available from the corresponding author upon reasonable request.

Acknowledgments: The author gratefully acknowledges DitchCarbon for providing firm-level greenhouse gas emissions data free of charge in support of the author's PhD research. DitchCarbon had no role in the design, analysis, interpretation, or publication of this study.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A

Appendix A.1

We begin by examining descriptive distribution diagnostics for all variables used in the empirical analysis, including excess returns, sustainability indicators, carbon exposure, controversy measures, macro-financial stress proxies, and transition-risk conditioning variables (EUA shocks and TRI). We report full-sample moments (mean, standard deviation, skewness, and kurtosis) and additionally examine annual subperiod summaries. This step is crucial in firm-level return panels because returns and sustainability measures are typically heavy-tailed and non-Gaussian. Documenting skewness and kurtosis motivates outlier-robust preprocessing and dependence-robust inference rather than reliance on normality assumptions. The year-by-year decomposition also allows assessment of whether extreme realizations cluster in specific periods, which is directly relevant to testing whether carbon pricing is state-dependent under systemic stress.

To reduce the risk of regression results being influenced by nonstationary behavior, we assess the time-series properties of key continuous variables using complementary panel unit root tests: the Levin–Lin–Chu (LLC) test, the Im–Pesaran–Shin (IPS) test, and a Fisher-type test based on augmented Dickey–Fuller statistics. Using multiple tests is deliberate because sustainability indicators can be persistent, tests impose different assumptions regarding homogeneity and cross-sectional structure, and a multi-test approach is more robust than relying on a single diagnostic.

Mechanical first differencing can remove economically meaningful low-frequency variation and complicate interpretation. Therefore, the workflow does not difference all firm-level characteristics by default. For slow-moving sustainability indicators, we construct stationarity-friendly alternatives while retaining levels for interpretability. Two complementary transformations are employed: (i) within-firm demeaning (the within transformation):

$$X_{i,t}^W = X_{i,t} - \bar{X}_i, \bar{X}_i = \frac{1}{T} \sum_t X_{i,t},$$

and (ii) firm-specific linear detrending:

$$X_{i,t} = a_i + b_i t + u_{i,t}, X_{i,t}^{DT} \equiv \hat{u}_{i,t}.$$

These transformations preserve within-firm variation (consistent with fixed-effects identification) and reduce the likelihood that estimated relationships are driven by slow drift in infrequently updated scores.

Carbon exposure is measured exclusively using the carbon score. Because the score is bounded and may exhibit mass points, we avoid nonlinear transforms that could behave poorly at the boundaries. Instead, we employ a rank-based annual percentile transformation:

$$C_{i,t}^{pct} = \frac{\text{rank}_{y(t)}(C_{i,t}) - 1}{N_{y(t)} - 1} \in [0, 1],$$

where $\text{rank}_{y(t)}(\cdot)$ denotes the within-year cross-sectional rank and $N_{y(t)}$ is the number of firms in year $y(t)$. This monotone mapping preserves the sample, reduces sensitivity to outliers and boundary effects, and is particularly suitable for interaction models with systemic stress.

Controversy is measured on an ordinal scale and is treated as a discrete event construct rather than a continuous process. We represent controversy using indicator variables such as a high-controversy state,

$$D_{i,t}^{High} = \mathbb{1}\{Level_{i,t} \geq 2\},$$

a severe-controversy state,

$$\mathbb{1}\{Level_{i,t} \geq 3\},$$

and an indicator capturing transitions into a high-controversy state,

$$D_{i,t}^{Up} = \mathbb{1}\{D_{i,t}^{High} = 1, D_{i,t-1}^{High} = 0\}.$$

This encoding aligns with the economic interpretation of controversies as discrete episodes and avoids imposing an artificial linear structure on ordinal categories. For robustness contexts where changes are preferable to levels, we also compute the quarterly change in the short-term risk-free proxy, $\Delta r_t^f = r_t^f - r_{t-1}^f$. After constructing these alternatives, stationarity diagnostics are rechecked to ensure that the transformations behave as intended.

Appendix A.2

The quarterly risk-free rate is derived from a short-term money market rate expressed as an annualized percentage and converted into a quarterly decimal rate:

$$r_t^f = \frac{\text{annualized short-term rate}_t}{100 \times 4}.$$

Firm-level excess returns are then defined as

$$r_{i,t}^{ex} = R_{i,t} - r_t^f,$$

where $R_{i,t}$ denotes the firm's quarterly total return. The return construction is carefully verified to ensure that conversions do not introduce missingness or systematic sample loss. This approach is standard in asset-pricing research: inference is conducted on risk-adjusted returns, and the explicit annual-to-quarter conversion prevents scaling inconsistencies that could distort coefficient magnitudes and interaction effects.

Appendix A.3

To improve comparability across firms and time while preserving the balanced panel, we apply structured outlier handling and scaling procedures.

First, continuous variables are winsorized at the 1st and 99th percentiles using Tukey-style winsorization:

$$X_{i,t}^W = \begin{cases} Q_{0.01}(X) & \text{if } X_{i,t} < Q_{0.01}(X), \\ X_{i,t} & \text{if } Q_{0.01}(X) \leq X_{i,t} \leq Q_{0.99}(X), \\ Q_{0.99}(X) & \text{if } X_{i,t} > Q_{0.99}(X). \end{cases}$$

Winsorization limits the leverage of extreme tail realizations without deleting observations. This is particularly relevant in stress-oriented return models, where extreme outcomes often coincide with economically meaningful crisis quarters.

Second, key firm-level characteristics are standardized using z scores within sector-time cells:

$$Z_{i,t}^{S \times t} = \frac{X_{i,t} - \mu_{s(i),t}(X)}{\sigma_{s(i),t}(X)}.$$

This sector–time standardization removes mechanical industry-level differences and expresses covariates in comparable units within homogeneous environments, ensuring that estimated carbon premia are not driven by static between-sector composition effects. For thin Sector $\times$ Quarter cells, we also compute a less granular Sector $\times$ Year standardization:

$$Z_{i,t}^{S \times Y} = \frac{X_{i,t} - \mu_{s(i),y(t)}(X)}{\sigma_{s(i),y(t)}(X)},$$

which is used in the core regressions to preserve sector neutrality while minimizing missingness. Macro/state variables that vary only over time and enter interaction terms (CISS, EUA shocks, and TRI) are standardized using global Z scores, $Z_t = (X_t - \mu(X))/\sigma(X)$, improving numerical stability in interaction models and supporting interpretation in “per one standard deviation” units.

Appendix B

Appendix B.1

Table A1 reports two-way fixed-effects regressions examining whether within-firm return sensitivity to carbon exposure varies with alternative transition-risk conditioning variables. Columns M4, M6, and M8 interact the standardized carbon-exposure measure with, respectively, systemic financial stress (CISS), EU ETS allowance price shocks (EUA), and a news-based Transition Risk Index (TRI), while controlling for ESG scores and controversy states. Because the conditioning variables vary only over time, their main effects are absorbed by quarter fixed effects, and identification relies exclusively on the interaction terms. Across specifications, the linear carbon coefficient remains economically small and statistically insignificant. The Carbon $\times$ EUA interaction is positive and marginally significant, whereas the Carbon $\times$ TRI interaction is statistically insignificant.

Table A1. Carbon–return interactions with transition-risk variables (continuous).

Variable	M4	M6	M8
	Carbon $\times$ CISS	Carbon $\times$ EUA	Carbon $\times$ TRI
CarbonScore_pct_st_SY	0.00128 (0.00328)	0.00129 (0.00327)	0.00119 (0.00328)
CISSz	0.03323 (0.01020)	—	—
Carbon $\times$ CISS	0.00310 (0.00233)	—	—
Carbon $\times$ EUA	—	0.00262 (0.00154)	—
Carbon $\times$ TRI	—	—	−0.00184 (0.00169)
ControvHigh	−0.00975 (0.00428)	−0.00957 (0.00428)	−0.00972 (0.00427)
ESGScore_st_SY	−0.00729 (0.00406)	−0.00747 (0.00406)	−0.00733 (0.00405)
Observations (firm–quarter)	6664	6664	6664
Groups	238	238	238
Within R ²	0.3245	0.3244	0.3242

Source: Author estimates from panel regressions estimated in Stata/MP 17.0 using the cleaned quarterly firm-level panel described in Table 2.

Because EUA shocks and TRI are standardized, the interaction coefficients can be interpreted in “per one standard deviation” units. The estimate for Carbon $\times$ EUA (0.00262) implies that a one-standard-deviation EUA shock is associated with approximately +0.26 percentage points higher quarterly excess returns for a one-standard-deviation-higher carbon exposure, whereas Carbon $\times$ TRI (−0.00184) implies approximately −0.18 percentage points (statistically insignificant). Overall, these magnitudes are economically modest.

Appendix B.2

Table A2 presents two-way fixed-effects regressions that permit return sensitivity to conditioning variables to vary nonparametrically across the carbon distribution. Carbon exposure is measured using quintile indicators, with the lowest-carbon group (Q1) serving as the reference category. Columns M5, M7, and M9 report interactions between carbon quintiles and, respectively, systemic financial stress (CISS), EU ETS allowance price shocks (EUA), and a news-based Transition Risk Index (TRI). Because the conditioning variables vary only over time, their main effects are absorbed by quarter fixed effects, and identification relies on the interaction terms. The quintile $\times$ CISS interaction terms are jointly significant. For transition-specific conditioning, the quintile $\times$ EUA block is jointly significant (F-test $p \approx 0.024$), whereas the quintile $\times$ TRI block is not jointly significant (F-test $p \approx 0.253$); neither pattern implies a stable or monotonic negative carbon premium in mean returns.

Table A2. Carbon–return interactions with transition-risk variables (continuous).

Variable	M3 Carbon Quintiles	M5 Quintiles CISS	M7 Quintiles EUA	M9 Quintiles TRI
CarbonScore_pct_st_SY	—	—	—	—
Q2	−0.01202 (0.00907)	−0.01145 (0.00904)	−0.01149 (0.00911)	−0.01215 (0.00910)
Q3	−0.02633 (0.01065)	−0.02554 (0.01061)	−0.02630 (0.01078)	−0.02637 (0.01068)
Q4	−0.01535 (0.01061)	−0.01509 (0.01059)	−0.01483 (0.01063)	−0.01583 (0.01061)
Q5	−0.00749 (0.01023)	−0.00702 (0.01022)	−0.00728 (0.01028)	−0.00785 (0.01026)
CISSz	—	0.02145 (0.01174)	—	—
Carbon $\times$ CISS	—	—	—	—
Q2 $\times$ CISS	—	0.02728 (0.00721)	—	—
Q3 $\times$ CISS	—	0.01101 (0.00667)	—	—
Q4 $\times$ CISS	—	0.00768 (0.00662)	—	—
Q5 $\times$ CISS	—	0.01451 (0.00711)	—	—
Carbon $\times$ EUA	—	—	—	—
Q1 $\times$ EUA (slope)	—	—	−0.00949 (0.00487)	—
Q2 $\times$ EUA (slope)	—	—	0.00324 (0.00537)	—
Q3 $\times$ EUA (slope)	—	—	−0.00664 (0.00573)	—
Q4 $\times$ EUA (slope)	—	—	0.00253 (0.00527)	—
Q5 $\times$ EUA (slope)	—	—	0 (omitted)	—
Carbon $\times$ TRI	—	—	—	—
Q1 $\times$ TRI (slope)	—	—	—	0.00014 (0.00526)
Q2 $\times$ TRI (slope)	—	—	—	−0.00281 (0.00550)
Q3 $\times$ TRI (slope)	—	—	—	−0.00180 (0.00483)
Q4 $\times$ TRI (slope)	—	—	—	−0.00949 (0.00507)
Q5 $\times$ TRI (slope)	—	—	—	0 (omitted)
ControvHigh	(included)	(included)	(included)	(included)
ESGScore_st_SY	(included)	(included)	(included)	(included)
Observations (firm–quarter)	6664	6664	6664	6664
Groups	238	238	238	238
Within R ²	0.3254	0.3294	0.3267	0.3260

Source: Author estimates from panel regressions estimated in Stata/MP 17.0 using the cleaned quarterly firm-level panel described in Table 2.

Appendix C

Appendix C.1 reports dispersion statistics for emissions and carbon proxies before and after transformation. Raw carbon exposure is concentrated at low values (CarbonScore ≤ 0.01 in 86.8% of observations; CarbonScore = 0 in 14.7%), motivating within-year percentile ranking and Sector $\times$ Year standardization. After transformation, CarbonScore_pct_st_SY is approximately standardized (SD ≈ 0.99 ; p5 ≈ -1.63 ; p95 ≈ 1.47) with limited tail mass (2.16% ≤ -2 SD; 1.20% $\geq +2$ SD). Appendix C.2 decomposes variation into within- and between-firm components; for CarbonScore_pct_st_SY, within SD ≈ 0.52 and between SD ≈ 0.85 , indicating that usable within-firm variation remains but is smaller than cross-sectional dispersion. Appendix C.3 reports two sensitivity checks: trimming extreme carbon observations (5% tails) and trimming extreme carbon firms (outside the 5th–95th percentile of firm-average carbon rank). In both cases, the linear carbon coefficient remains economically small and statistically indistinguishable from zero, supporting the conclusion that the baseline null is not driven by a small number of tail observations within the ESG Leaders universe. Appendix C.4 reports alternative binning using carbon quintiles constructed within Sector $\times$ Year cells and the corresponding joint test of the quintile block, alongside the baseline global-quintile definition for comparison. Appendix C.5 reports an industry-time-shock robustness check replacing quarter fixed effects with Sector $\times$ Quarter fixed effects (firm FE + Sector $\times$ Quarter FE); the linear carbon coefficient remains economically small and statistically indistinguishable from zero, indicating that the baseline null is not explained by time-varying industry shocks.

Appendix C.1

Table A3 reports dispersion statistics for emissions and carbon proxies before and after transformation. CarbonScore is a bounded raw proxy constructed from total GHG emissions. CarbonScore_pct is the within-year percentile rank of CarbonScore, and CarbonScore_pct_st_SY is the Sector $\times$ Year standardized version used in the regressions. Location statistics across rows are therefore not directly comparable.

Table A3. Dispersion statistics for emissions and carbon exposure measures.

Variable	Observations (Firm–Quarter)	Mean	SD	p5	p25	p50	p75	p95	IQR
ln_GHG_Total	6664	14.20	7.39	0.00	10.54	15.61	19.47	24.55	8.93
CarbonScore	6664	0.018	0.085	0.000	0.001	0.014	0.025	0.111	0.025
CarbonScore_pct	6664	0.483	0.271	0.046	0.248	0.498	0.707	0.896	0.459
CarbonScore_pct_st_SY	6664	0.000	0.9940	−1.6320	−0.765	0.028	0.793	1.475	1.558

Source: Author calculations based on the quarterly firm-level panel described in Table 2.

Appendix C.2

Table A4 decomposes variation into within- and between-firm components. For CarbonScore_pct_st_SY, within SD ≈ 0.52 and between SD ≈ 0.85 , indicating that nontrivial within-firm variation remains for fixed-effects identification.

Table A4. Within–between decomposition of carbon exposure.

Variable	Overall SD	Between SD	Within SD	Within/ Overall Variance
CarbonScore	0.0850	0.0838	0.0151	0.03
CarbonScore_pct	0.2710	0.2382	0.1302	0.23
CarbonScore_pct_st_SY	0.9943	0.8517	0.5159	0.27

Source: Author calculations based on the quarterly firm-level panel described in Table 2.

Appendix C.3

Table A5 reports two sensitivity checks: trimming extreme carbon observations (5% tails) and trimming extreme carbon firms (outside the 5th–95th percentile of firm-average carbon rank). Across both trims, the linear carbon coefficient remains economically small and statistically indistinguishable from zero.

Table A5. Range-restriction sensitivity: trimming extremes.

Specification	Carbon Coefficient	Std. Error	<i>p</i> -Value	Observations (Firm–Quarter)
Baseline FE	0.00124	0.00328	0.706	6664
Trim obs. (5% tails)	0.00167	0.00333	0.616	6284
Trim firms (5–95% mean)	0.00156	0.00329	0.636	6076

Source: Author calculations based on the quarterly firm-level panel described in Table 2.

Appendix C.4

Table A6 reports alternative binning using carbon quintiles constructed within Sector $\times$ Year cells. Under this alternative binning, the quintile block is jointly significant ($p = 0.0299$), whereas the corresponding joint test under global quintiles is marginal ($p \approx 0.057$).

Table A6. Alternative binning (carbon quintiles within Sector $\times$ Year).

Quintile Definition	Q2	Q3	Q4	Q5	Joint Test of Quintile Block
Global quintiles (CarbonQ)	−0.0120 (0.0091)	−0.0263 (0.0107)	−0.0154 (0.0106)	−0.0075 (0.0102)	$p \approx 0.057$
Sector $\times$ Year quintiles (CarbonQ_SY)	0.0021 (0.0077)	−0.0168 (0.0089)	−0.0067 (0.0094)	−0.0022 (0.0092)	$p = 0.0299$

Source: Author calculations based on the quarterly firm-level panel described in Table 2.

Appendix C.5

Table A7 reports industry time shocks: replacing quarter fixed effects with Sector $\times$ Quarter fixed effects. Dependent variable: quarterly excess return (ret_excess). All specifications include firm fixed effects and firm-clustered standard errors. Column (2) replaces quarter fixed effects with Sector $\times$ Quarter fixed effects (sec_q) to absorb time-varying industry shocks. In both specifications, the linear carbon coefficient remains economically small and statistically indistinguishable from zero (baseline $p = 0.7060$; Sector $\times$ Quarter $p = 0.7160$), indicating that the null linear result is not explained by sector-level time-varying shocks.

Table A7. Sector $\times$ Quarter fixed-effects robustness.

Quintile Definition	CarbonScore_ pct_st_SY	ESGScore_st_SY	Controv High	Observations (Firm–Quarter)
Firm FE + Quarter FE (baseline)	0.0012 (0.0033)	−0.0074 (0.0041)	−0.0096 (0.0043)	6664
Firm FE + Sector $\times$ Quarter FE	0.0012 (0.0033)	−0.0071 (0.0039)	−0.0057 (0.0043)	6664

Source: Author calculations based on the quarterly firm-level panel described in Table 2.

Notes

- ¹ In this paper, carbon risk denotes a firm's exposure to transition-related financial risk, proxied by observable greenhouse gas emissions. We use total reported emissions as an aggregate indicator, consistent with how market participants typically observe and process carbon information.
- ² We employ aggregate emissions rather than Scope 1, 2, or 3 measures because scope-level reporting varies significantly across firms and over time, particularly for Scope 3, introducing avoidable measurement noise.
- ³ In this context, measurement noise refers to variation in ESG scores arising from differences in data availability, indicator selection, weighting schemes, and aggregation methods across rating providers, rather than from actual differences in firms' environmental exposure. Consequently, observed ESG scores may diverge from the underlying sustainability characteristics they intend to capture, reducing their reliability as proxies for risk-relevant fundamentals (Berg et al. 2022).
- ⁴ Firm selection from the MSCI Europe universe is benchmark-driven rather than ad hoc. MSCI Europe represents Developed Markets Europe and covers large- and mid-cap equities across fifteen developed European markets (Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom), accounting for approximately 85% of the free-float-adjusted market capitalization of the European developed equity universe. The index emphasizes size, liquidity, and free float to ensure investability and replicability for institutional portfolios. This focus is particularly relevant for carbon-risk analysis because carbon exposure is primarily priced through institutional rebalancing, benchmark tracking, and risk-management channels that operate most effectively in large, liquid markets. Including smaller or less investable markets would introduce thin trading, episodic price discovery, and heterogeneous disclosure practices, weakening the identification of carbon–return relationships.
- ⁵ Expanding the sample beyond the MSCI Europe and MSCI ESG Leaders universes would substantially reduce the availability and consistency of historical ESG, carbon, and controversy data, particularly in the early years when corporate disclosure and third-party coverage were limited. Emissions and controversy data remain uneven across European firms, and broader coverage would increase missing or unreliable observations, undermining the construction of a balanced quarterly panel. To avoid inconsistencies in Scope-level reporting, especially in earlier periods, the analysis relies on firms' reported total GHG emissions rather than aggregating Scope 1–3 components. Data providers such as Sustainalytics, MSCI, and DitchCarbon also exhibit incomplete archival coverage outside core European indices, and pay-per-point pricing structures impose practical constraints on large-scale historical expansion. Restricting the sample to firms with consistent coverage reflects a data-driven choice that preserves longitudinal comparability and avoids ad hoc reconstruction.
- ⁶ Reproducibility and ethical considerations. Data inputs include public series (ECB Statistical Data Warehouse), licensed sources (S&P Capital IQ Pro, MSCI ESG Research, Sustainalytics, and DitchCarbon), and hand-collected information from corporate sustainability reports and regulatory disclosures. As several inputs are proprietary, raw vendor data cannot be redistributed. To facilitate replication, all variable constructions, timing and alignment rules, and estimation procedures are documented, and an audit trail is maintained for hand-collected observations (firm–year sources and definitions). The study does not involve human subjects or personal data. Therefore, ethical approval and informed consent are not required.
- ⁷ S&P Capital IQ Pro is an analytics platform from S&P Global that provides access to a wide range of financial, market, and ESG data, including MSCI ratings, asset returns, market indices, and macroeconomic variables. <https://www.spglobal.com/market-intelligence/en/solutions/products/sp-capital-iq-pro> (accessed on 30 August 2025).
- ⁸ European Central Bank. ECB Statistical Data Warehouse. <https://data.ecb.europa.eu/data/datasets> (accessed on 3 October 2025).
- ⁹ Sustainalytics is a leading ESG analytics company owned by Morningstar. It provides ESG risk ratings and assessments for companies worldwide. <https://www.sustainalytics.com/> (accessed on 30 August 2025).

- 10 For firms with missing or incomplete emissions data, a static carbon proxy, such as the DitchCarbon rating (accessed on 29 April 2025 and 20 December 2025), may be used to preserve cross-sectional coverage. For firms with partial carbon reporting histories, missing intermediate years were imputed using Huber regression, which provides robust trend estimation resistant to outliers while maintaining sensitivity to underlying emission trajectories. This method ensures a continuous and comparably scaled carbon score series across the entire sample period.
- 11 Due to inconsistent availability of emissions intensity data, the analysis uses absolute emission values and applies cross-sectional normalization within each year to construct a comparable carbon score ranging from 0 to 1. The score reflects the relative carbon footprint of firm i in year t , where higher scores indicate higher emissions relative to industry peers.
- 12 Several alternative transition-risk information measures have been introduced in the literature, such as the negative climate change news indices developed by Engle et al. (2020) and subsequently applied by Ilhan et al. (2021). These measures are generally constructed at a monthly frequency and predominantly utilize U.S.-based news sources or search-intensity data, incorporating additional filtering procedures such as innovation extraction. Such approaches are not directly compatible with the quarterly European panel structure employed in this study. In contrast, EU ETS allowance price shocks and the Transition Risk Index are inherently suited to quarterly frequency and the European institutional context, rendering them more appropriate for the current analysis.
- 13 We use quarterly changes in EU ETS allowance prices (EUA) as a transition-salience proxy because the EU ETS is the EU's central market-based decarbonization instrument, where one allowance confers the right to emit one tonne of CO₂-equivalent and the allowance price summarizes the marginal cost of regulated emissions (European Commission n.d.). Importantly, EUA prices have been shown to respond to changes in EU climate-policy stringency and market-design reforms (including the Green Deal/Fit-for-55 context), making them a plausible high-frequency market-based indicator of tightening (or easing) transition-cost pressure (Ampudia et al. 2022). Methodologically, carbon-market price changes around regulatory updates are also used in the literature to identify carbon-policy surprises, supporting the interpretation of EUA price innovations as transition-policy shocks rather than slow-moving trends (Känzig 2021). EU ETS allowance prices (data vendor/exchange) <https://tradingeconomics.com/commodity/carbon> (accessed on 29 January 2026).
- 14 TRI is a text-based index constructed from Reuters news to capture time variation in transition-risk attention (policy, regulation, and technology-related transition developments). It is designed to spike when transition-related discussion intensifies and is therefore a direct proxy for innovations in the public transition-risk information environment. We use TRI aggregated to quarterly frequency to match the return panel (Bua et al. 2022). Transition Risk Index (TRI) datasets are also available at <http://www.policyuncertainty.com> (accessed on 29 January 2026).
- 15 In the Table 3 the variable suffixes indicate preprocessing steps described in Section 4. For carbon exposure, CarbonScore_pct is computed as a within-year cross-sectional percentile rank of CarbonScore, $(\text{rank}_{\text{Year}}(\text{CarbonScore}) - 1) / (N_{\text{Year}} - 1)$, mapping observations onto $[0, 1]$. The suffix st_SY indicates Sector $\times$ Year z normalization: for each Sector $\times$ Year cell, the variable is centered by the cell mean and scaled by the cell standard deviation. Thus, CarbonScore_pct_st_SY is the Sector $\times$ Year standardized percentile-rank carbon score, and CarbonScore_pct_st_SY² is its squared term. ESGScore_st_SY is the Sector $\times$ Year standardized ESG composite score computed analogously. The suffix z indicates a global z score over the pooled sample (e.g., CISSz). The categorical variable CarbonQ is formed as quintiles of CarbonScore_pct_st_SY (Q1 is the reference group), and rows Q2–Q5 in the tables report the corresponding indicator coefficients. ControvHigh is a binary indicator defined as $1\{\text{"ESGControversyLevel"} \geq 2\}$. Interaction rows (e.g., Carbon $\times$ CISS and Qk $\times$ CISS) correspond to the relevant interaction terms with CISSz. Versions 4 and 5 of the output report one omitted quarter dummy due to collinearity. ControvHigh is defined as $1\{\text{"ESGControversyLevel"} \geq 2\}$. CarbonScore_pct_st_SY and ESGScore_st_SY are standardized as described in Section 4.
- 16 Marginal effects of systemic stress are computed from the carbon quintile interaction model in Table 3 (Column 5). Because the specification is linear in CISSz, the marginal effect is the quintile-specific slope with respect to stress: for the reference group Q1, $\partial \mathbb{E}[\text{Return}_{i,t}] / \partial \text{CISS}_{t,z} = \beta_{\text{CISS}}$; for Qk ($k = 2, \dots, 5$), $\partial \mathbb{E}[\text{Return}_{i,t}] / \partial \text{CISS}_{t,z} = \beta_{\text{CISS}} + \beta_{\text{Qk} \times \text{CISS}}$. Standard errors and 95% confidence intervals for these linear combinations are obtained using the estimated variance–covariance matrix (delta method). Because CISSz is standardized, marginal effects represent the change in quarterly returns associated with a one-standard-deviation increase in systemic stress, conditional on firm and quarter fixed effects.
- 17 Driscoll–Kraay and alternative clustering change the variance estimator; point estimates are identical to Table 3. Lagged-regressor models reduce the sample size by one quarter. Carbon_mean and ESG_mean are firm-level averages of the standardized variables used to implement correlated random effects (Mundlak).

References

- Alessi, Lucia, Elisa Ossola, and Roberto Panzica. 2023. When do investors go green? Evidence from a time-varying asset-pricing model. *International Review of Financial Analysis* 90: 102898. [CrossRef]
- Al Rabab'a, Eltayyeb Al-Fakir, Afzalur Rashid, Syed Shams, and Sudipta Bose. 2024. Corporate carbon performance and firm risk: Evidence from Asia-Pacific countries. *Journal of Contemporary Accounting & Economics* 20: 100427. [CrossRef]

- Ampudia, Miguel, Giovanna Bua, Daniel Kapp, and Dilyara Salakhova. 2022. The Role of Speculation during the Recent Increase in EU Emissions Allowance Prices. *ECB Economic Bulletin* 3: 58–63.
- Aswani, Jitendra, Aneesh Raghunandan, and Shiva Rajgopal. 2024. Are carbon emissions associated with stock returns? *Review of Finance* 28: 75–106. [\[CrossRef\]](#)
- Avramov, Doron, Si Cheng, Abraham Lioui, and Andrea Tarelli. 2022. Sustainable investing under ESG rating uncertainty. *Journal of Financial Economics* 145: 642–64. [\[CrossRef\]](#)
- Bang, Jeongseok, Doojin Ryu, and Robert I. Webb. 2023. ESG controversy as a potential asset-pricing factor. *Finance Research Letters* 58: 104315. [\[CrossRef\]](#)
- Berg, Florian, Julian F. Kölbel, and Roberto Rigobon. 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance* 26: 1315–44. [\[CrossRef\]](#)
- Bolton, Patrick, and Marcin T. Kacperczyk. 2021. Do investors care about carbon risk? *Journal of Financial Economics* 142: 517–49. [\[CrossRef\]](#)
- Bolton, Patrick, and Marcin T. Kacperczyk. 2023. Global pricing of carbon-transition risk. *The Journal of Finance* 78: 3677–724. [\[CrossRef\]](#)
- Bose, Sudipta, Edwin KiaYang Lim, Kristina Minnick, Patrick J. Schorno, and Syed Shams. 2025. Does carbon risk influence stock price crash risk? International evidence. *Journal of Business Finance & Accounting* 52: 2111–44. [\[CrossRef\]](#)
- Boubaker, Sabri, Tonmoy Choudhury, Fakhru Hasan, and Duc Khuong Nguyen. 2024. Firm carbon risk exposure, stock returns, and dividend payment. *Journal of Economic Behavior & Organization* 221: 248–76. [\[CrossRef\]](#)
- Bua, Giovanna., Daniel Kapp, Federico Ramella, and Lavinia Rognone. 2022. Transition Versus Physical Climate Risk Pricing in European Financial Markets: A Text-Based Approach. ECB Working Paper No. 2677. Available online: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp26779fc49e8300.en.pdf> (accessed on 16 December 2025).
- Bui, Binh, Muhammad Nurul Houqe, and Muhammad Kaleem Zahir-ul-Hassan. 2022. Moderating effect of carbon accounting systems on strategy and carbon performance: A CDP analysis. *Journal of Management Control* 33: 483–524. [\[CrossRef\]](#)
- 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: 147–75. [\[CrossRef\]](#)
- Deng, Jialu, Wenhe Lin, Jiuzhen Huang, Yunying Cai, and Wenbao Wang. 2025. Research on the impact of the ESG rating divergence of manufacturing firms on carbon emission intensity. *PLoS ONE* 20: E0323929. [\[CrossRef\]](#)
- Dong, Yao, Martina Hengge, Fabián Valencia, and Richard Varghese. 2025. *Carbon Risk in Loan Pricing*. IMF Working Paper No. WP/25/250. Washington, DC: International Monetary Fund.
- Duppati, Geeta. 2025. Carbon pricing and revenue growth valuation risk exposure. *International Review of Economics & Finance* 103: 104442. [\[CrossRef\]](#)
- Ehlers, Torsten, Frank Packer, and Kathrin de Greiff. 2021. *The Pricing of Carbon Risk in Syndicated Loans: Which Risks Are Priced and Why?* BIS Working Papers No. 946. Basel: Bank for International Settlements.
- Engle, Robert F., Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel. 2020. Hedging climate change news. *Review of Financial Studies* 33: 1184–216. [\[CrossRef\]](#)
- European Commission. n.d. *About the EU ETS*. Brussels: Directorate-General for Climate Action. Available online: https://climate.ec.europa.eu/eu-action/carbon-markets/about-eu-ets_en (accessed on 29 January 2026).
- Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [\[CrossRef\]](#)
- Holló, Dániel, Manfred Kremer, and Marco Lo Duca. 2012. *CISS—A Composite Indicator of Systemic Stress in the Financial System*. ECB Working Paper Series No. 1426 (March); Frankfurt am Main: European Central Bank. Available online: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1426.pdf> (accessed on 30 December 2025).
- Hou, Kewei, Chen Xue, and Lu Zhang. 2015. Digesting anomalies: An investment approach. *The Review of Financial Studies* 28: 650–705. [\[CrossRef\]](#)
- Ilhan, Emirhan, Zacharias Sautner, and Grigory Vilkov. 2021. Carbon tail risk. *Review of Financial Studies* 34: 1540–71. [\[CrossRef\]](#)
- Känzig, Diego R. 2021. *The Economic Consequences of Putting a Price on Carbon*. Working paper. London: London Business School. Available online: https://www.ecb.europa.eu/press/conferences/ecbforum/shared/pdf/2021/kaenzig_paper.bg.pdf (accessed on 30 December 2025).
- Li, Jinpei, Jiaxin Suo, Yixuan Huang, and Xiangdong Sun. 2025. The carbon reduction effect of ESG performance: Empirical evidence from Chinese shipping enterprises. *Frontiers in Marine Science* 12: 1575398. [\[CrossRef\]](#)
- Li, Shiyu, Xiangyuan Ao, Mengdie Zhang, and Mingyue Pu. 2024. ESG performance and carbon emission intensity: Examining the role of climate policy uncertainty and the digital economy in China’s dual-carbon era. *Frontiers in Environmental Science* 12: 1526681. [\[CrossRef\]](#)
- Luo, Qin, and Feng Ma. 2025. Carbon transition risk and stock market premium. *International Review of Financial Analysis* 105: 104458. [\[CrossRef\]](#)

- Monasterolo, Irene, and Luca de Angelis. 2020. Blind to carbon risk? An analysis of stock market reaction to the Paris Agreement. *Ecological Economics* 170: 106571. [\[CrossRef\]](#)
- Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor. 2021. Sustainable investing in equilibrium. *Journal of Financial Economics* 142: 550–71. [\[CrossRef\]](#)
- Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski. 2021. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142: 572–97. [\[CrossRef\]](#)
- Saha, Rupjyoti, and Santi Gopal Maji. 2025. Tangibility of intangibles in the ESG disclosure–carbon intensity nexus for S&P 500 firms. *Managerial Finance* 51: 1377–91. [\[CrossRef\]](#)
- Serafeim, George, and Aaron Yoon. 2022. Which corporate ESG news does the market react to? *Financial Analysts Journal* 78: 59–78. [\[CrossRef\]](#)
- Xie, Hanjin, Zilong Qin, and Jun Li. 2024. ESG performance and corporate carbon emission intensity: Based on panel data analysis of A-share listed companies. *Frontiers in Environmental Science* 12: 1483237. [\[CrossRef\]](#)
- Yang, Yongliang, Jitao Zhang, Changting Song, Yingying Wu, and Bingbing Zhang. 2025. Does environmental information disclosure make financial markets more informative? Evidence from China. *Journal of Contemporary Accounting & Economics* 21: 100468. [\[CrossRef\]](#)
- Ye, Jian, and Wenhua Xu. 2023. Carbon reduction effect of ESG: Empirical evidence from listed manufacturing companies in China. *Frontiers in Ecology and Evolution* 11: 1311777. [\[CrossRef\]](#)
- Zhou, Ye, Yige Fan, Siyan Shen, and Zeyun Bei. 2025. The carbon risk premium: Evidence from Chinese stock market. *Asia-Pacific Journal of Accounting & Economics*, 1–24. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.