

TALK, WALK, AND THE CSRD

EVIDENCE ON THE DISCLOSURE-PERFORMANCE GAP

GABRIEL FANNBERG

RYAN PAGE

Bachelor Thesis

Stockholm School of Economics

2026



Talk, Walk, and the CSRD: Evidence on the Disclosure-Performance Gap

Abstract:

Whether mandatory sustainability disclosure reflects real corporate performance is central to evaluating the effectiveness of recent regulation. This paper studies whether the first year of the EU's Corporate Sustainability Reporting Directive narrowed the gap between firms' sustainability disclosure ("talk") and third-party-assessed performance ("walk"). We exploit the directive's 500-employee scope threshold in a difference-in-differences design on 132 European listed firms observed from 2021 to 2024. The composite talk-walk gap widens after CSRD, and the widening is concentrated in the environmental dimension. A decomposition shows that the change reflects a surge in disclosure language with no movement in assessed performance. The effect is largest among firms with the most disclosure ground to cover, namely non-Nordic firms and firms whose pre-period talk and walk were already aligned. This pattern is consistent with compliance-driven expansion of mandated reporting rather than deliberate overclaiming, and highlights a measurement issue that any text-based study of greenwashing in regulated settings will need to confront.

Keywords:

CSRD; ESRS; Sustainability disclosure; text-as-data; ESG performance

Authors:

Gabriel Fannberg (26079)

Ryan Page (25973)

Tutor:

Ye Zhang, Assistant Professor, Department of Finance

Examiner:

Ramin Baghai, Professor, Department of Finance

Bachelor Thesis

Bachelor Program in Business and Economics

Stockholm School of Economics

© Gabriel Fannberg and Ryan Page, 2026

1. Introduction

Mandatory sustainability disclosure has become one of the most ambitious instruments of European corporate regulation in the past decade. The EU’s Corporate Sustainability Reporting Directive (CSRD), adopted in December 2022, requires in-scope firms to report under the European Sustainability Reporting Standards (ESRS) and to subject those disclosures to third-party assurance. The directive’s stated aim is to make sustainability information comparable, verifiable, and decision-useful for investors, regulators, and other stakeholders. Whether the directive achieves that aim depends on a basic question: does mandated reporting bring firms’ sustainability claims closer to their actual performance, or does it primarily expand the volume of language without changing what firms do? This paper studies whether Wave 1 of CSRD narrowed the gap between firms’ sustainability disclosure (“talk”) and third-party-assessed performance (“walk”).

Our empirical strategy uses a difference-in-differences design on 132 European listed firms observed from 2021 to 2024, exploiting the directive’s 500-employee scope threshold. The estimation sample is restricted to firms with 125 to 1,000 employees, placing treated and control firms in a comparable size range while preserving statistical power. Talk is measured from the text of annual reports using a keyword dictionary built on [Andrikogiannopoulou, Krueger, Mitali, and Papakonstantinou \(2022\)](#) and extended with CSRD-specific vocabulary. Walk is measured using LSEG’s ESG pillar scores, which aggregate verified emissions, resource-use, and innovation indicators rather than relying solely on firm self-reporting ([Aswani, Raghunandan, and Rajgopal, 2024](#)). Both measures are standardised as z -scores using pre-period (2021–2023) means, so the gap is expressed in units of pre-period standard deviations.

The talk–walk gap widened, not narrowed, in CSRD’s first reporting year. The composite gap rises by roughly two-thirds of a pre-period standard deviation for treated firms relative to controls, and the widening is concentrated in the environmental dimension where ESRS imposes its most detailed requirements. A decomposition shows that the entire widening comes from talk: treated firms’ sustainability language rose sharply after CSRD, while their assessed environmental performance did not move. A dynamic version of the design places almost all of the talk adjustment at 2024, the first year of mandatory ESRS reporting, and shows no differential pre-trend in the immediate pre-period. The findings are robust to alternative bandwidths, the choice of controls, winsorisation, the exclusion of 2021, restriction to a balanced panel, and placebo timing.

The direction of the effect was not obvious in advance. CSRD pulls on two channels that push in opposite directions. The first is a disciplining channel: by requiring standardised reporting and external assurance, the directive raises the cost of unsubstantiated claims. A firm that previously embedded vague commitments in a bespoke sustainability section now faces a structured template in which omissions and inconsistencies are visible to investors, auditors, and regulators. Under this channel, firms inflating their rhetoric relative to their performance can scale back the talk or improve the walk; both responses narrow the gap. This is the prediction of theories of mandatory disclosure such as [Admati and Pfleiderer \(2000\)](#), in which forced disclosure improves market efficiency by constraining strategic withholding.

The second is a compliance channel. ESRS does not merely require firms to substan-

tiate existing claims; it compels them to report across a comprehensive set of topics, including environmental policies, emission targets, climate risk assessments, and transition plans, regardless of whether the firm has made meaningful progress in those areas. This mechanically inflates the textual footprint of sustainability disclosure. Environmental performance is sticky, reflecting physical investments and verified emissions reductions that take years to materialise, whereas sustainability language can be updated within a single reporting cycle. If the compliance channel dominates, the gap widens in the first reporting year not because firms are greenwashing more deliberately, but because the regulation itself mandates new vocabulary.

Our heterogeneity evidence is consistent with the compliance channel. The effect is largest for non-Nordic firms, which had less developed voluntary reporting traditions before CSRD, and for firms whose pre-period talk and walk were already aligned, which had the most disclosure ground to cover under ESRS. Deliberate overclaiming would predict the opposite pattern, with the largest response among firms with the highest pre-period gaps. Our design cannot cleanly separate compliance from deliberate overclaiming, however, and we therefore frame our findings as a disclosure–performance gap rather than greenwashing in a normative sense.

Our findings speak directly to recent work on the design of sustainability regulation. [Inderst and Opp \(2025\)](#) argue that a mandatory taxonomy is necessary to discipline greenwashing in ESG product markets; their analysis assumes that forced disclosure constrains cheap talk. Our evidence suggests that any disciplining effect is slow to appear: in CSRD’s first reporting year, treated firms’ sustainability language expands sharply while their assessed performance does not move. We contribute to this literature by providing the first firm-level test of CSRD’s effect on a disclosure–performance gap, and by flagging a measurement issue that any text-based study of greenwashing in regulated settings will need to confront: when a regime mandates vocabulary, dictionary-based measures of corporate language cannot cleanly separate substantive claims from compliance.

Related Literature

Our paper first contributes to a theoretical literature on how mandatory disclosure regulation shapes corporate communication. Existing work shows that forced disclosure improves market efficiency in settings where firms would otherwise strategically withhold or distort information ([Admati and Pfleiderer, 2000](#)), and [Inderst and Opp \(2025\)](#) extend this logic to ESG product markets, arguing that a mandatory taxonomy is necessary to discipline greenwashing because voluntary disclosure does not constrain cheap talk. These models implicitly treat forced disclosure as a tightening device: firms compelled to substantiate claims have less room to inflate them. Unlike these theoretical studies, we provide direct empirical evidence on how a binding disclosure mandate affects the gap between firm rhetoric and firm action. We document that, in its first reporting year, CSRD widens rather than narrows this gap, and our heterogeneity results trace the widening to a compliance channel rather than to deliberate overclaiming. Our findings qualify the timing implicit in these models. Talk responds within a single reporting cycle because firms can update language quickly; walk is sticky because environmental performance reflects physical investments that take longer to materialise. Whether the disciplining effect eventually appears once walk has had time to respond is something our

single post-period cannot establish.

Our paper is also directly related to a growing literature that uses text-based methods to measure the content of corporate communication. [Hoberg and Phillips \(2016\)](#) establish text-based methods for measuring product differentiation; [Sautner, van Lent, Vilkov, and Zhang \(2023\)](#) construct firm-level climate exposure measures from earnings call transcripts using a keyword approach; and [Andrikogiannopoulou et al. \(2022\)](#) build an ESG keyword library from mutual fund prospectuses. These papers consistently emphasise that dictionary-based measures capture attention to a topic rather than the substantive content or credibility of what firms say, and [Sautner et al. \(2023\)](#) label their measure climate-change *exposure* for precisely this reason. Complementing these studies, we extend the [Andrikogiannopoulou et al. \(2022\)](#) library with CSRD-specific vocabulary and apply it to firm annual reports, treating the resulting measure as one of disclosure intensity rather than commitment quality. We also flag a measurement issue with broader relevance: when a disclosure regime mandates vocabulary, dictionary-based text measures cannot cleanly separate substantive claims from mandated compliance. Any future text-based study of greenwashing in regulated settings will need to confront this problem, either through claims-level natural-language extraction or through a longer post-period in which performance has had time to respond.

Lastly, we contribute to a literature on the measurement, pricing, and real effects of ESG performance. [Aswani, Raghunandan, and Rajgopal \(2024\)](#) document that raw firm-reported emissions data suffer from inconsistent reporting boundaries and estimation error; [Bolton and Kacperczyk \(2021\)](#) and [Ilhan, Sautner, and Vilkov \(2021\)](#) find that carbon risk is priced in equity markets; [Krueger, Sautner, and Starks \(2020\)](#) document that institutional investors care about climate risk; [Pástor, Stambaugh, and Taylor \(2021\)](#) embed sustainable investing in an equilibrium asset-pricing framework; and [Hong and Kacperczyk \(2009\)](#) and [Duchin, Gao, and Xu \(2025\)](#) study how social norms and pollution exposure interact with corporate decisions and asset markets. These papers establish that ESG performance varies across firms in ways that the market prices, and that measurement choices materially affect inference. Building on these insights, we use third-party performance measures validated in this literature rather than self-reported emissions, and report cross-sectional evidence on how the talk-walk gap covaries with firm size and country of incorporation.

Most directly, we provide what is, to our knowledge, the first firm-level test of CSRD’s effect on a disclosure-performance gap. Existing work either evaluates voluntary disclosure regimes and infers misalignment from the cross section of self-reported metrics, or studies CSRD descriptively. Unlike these existing studies, we identify a causal effect using the directive’s sharp employee threshold and a within-firm difference-in-differences design that separates disclosure intensity from third-party-assessed performance using independent data sources. Whether the disciplining effect anticipated by [Inderst and Opp \(2025\)](#) eventually materialises, narrowing the gap as walk catches up to talk, is the central question we leave open to future research.

Having situated our contribution relative to the existing literature, we now outline the structure of the paper. The remainder of the paper is organised as follows. Section 2 describes the institutional setting. Section 3 describes the data and the construction of our talk, walk, and gap measures. Section 4 sets out the empirical strategy and presents the

main results. Section 5 reports cross-sectional evidence from a broader universe sample. Section 6 presents robustness checks. Section 7 discusses the findings and Section 8 concludes.

2. Institutional Background

The Corporate Sustainability Reporting Directive was adopted by the European Parliament and Council in December 2022 (European Commission, 2022). CSRD strengthens and extends the earlier Non-Financial Reporting Directive (NFRD) and requires in-scope firms to disclose, first, how sustainability issues create risks and opportunities for the business, and second, how the firm’s own activities affect people and the environment. This is the so-called double-materiality principle. A central element is standardisation: firms subject to CSRD must report under the European Sustainability Reporting Standards, which impose a common template across environmental, social, and governance topics and are subject to third-party assurance.

Implementation is phased. Wave 1 applies to firms that were already subject to NFRD reporting, namely large public-interest entities, and requires reporting for the 2024 financial year, with reports published in 2025. Under the NFRD, the scope is tied to public-interest entities with more than 500 employees, including listed companies, banks, and insurers. This pre-existing scope rule inherits directly into CSRD Wave 1 and provides the threshold-based treatment assignment we exploit in Section 4. Subsequent waves were intended to expand coverage to smaller listed firms and non-listed undertakings, but the EU has adopted a “stop-the-clock” delay that effectively pushed these back, leaving Wave 1 as the main binding cohort in our near-term window. Figure 1 summarises the adoption timeline and Figure 2 illustrates the threshold-based identification setting.

Conceptually, CSRD and ESRS raise the cost of “cheap talk” by making sustainability claims easier to verify and compare. The directive could, in principle, nudge firms to temper exaggerated language, improve underlying performance, or both. These are precisely the channels our talk–walk misalignment measures are designed to capture.

3. Data and Measurement

The analysis uses two samples. The first is a universe sample of European listed firms covered by LSEG’s ESG data, used for descriptive and cross-sectional evidence in Section 5. The second is the difference-in-differences estimation sample, constructed around the 500-employee CSRD threshold and used for the main analysis in Section 4. We describe each in turn and then the construction of talk, walk, and the gap. Implementation details, including the full keyword dictionary, weighting scheme, and LSEG variable definitions, are provided in Appendix A and Appendix B.

3.1 Data sources

We draw on two primary sources: structured data from LSEG Workspace, and firm-level annual reports. LSEG Workspace, operated by the London Stock Exchange Group, provides comprehensive data on firm financials, corporate documentation, and sustainability

metrics. We use LSEG’s ESG pillar scores, emissions data, and policy-level indicators. LSEG is widely used in the academic finance literature, and its ESG coverage incorporates verified emissions data, resource-use metrics, and third-party environmental innovation assessments rather than relying solely on firm self-reporting (Aswani et al., 2024). For the DiD sample we additionally collect English-language annual reports for the full analysis period (financial years 2021 through 2024), from which we extract the text-based talk measure described below.

3.2 Sample construction

Universe sample. We apply an initial equity screen in LSEG Workspace to isolate European listed firms, yielding 1,475 candidates. We retain firms in LSEG’s “Clean” subset (those with non-missing ESG pillar scores across the analysis period), reducing the sample to 996 firms. Excluding 15 firms with more than one missing year for the core regression variables yields a final universe panel of 981 firms over four years and 3,716 firm-year observations. We exclude 2025 because ESG data and annual report publication for FY2025 remain incomplete at the time of analysis. Table 1 reports the step-by-step attrition. Because the universe sample contains essentially no firms below 500 employees (the minimum is 504), it cannot replicate the treated-control comparison underlying the DiD; we use it for descriptive and correlational analysis only.

DiD estimation sample. We apply an equity screen to identify European-domiciled listed firms in the 125–1,000 employee bandwidth around the CSRD threshold, yielding an initial sample of 232 firms. For each firm we collect English-language annual reports for financial years 2021 through 2024 where available. After processing these reports through the talk extraction pipeline, we obtain talk scores for 208 firms; the remaining 24 either did not publish English-language reports or yielded reports from which text extraction failed. Merging talk scores with LSEG walk data and firm metadata leaves 205 firms, with three lost to identifier mismatches. Re-imposing the employee bandwidth on LSEG’s employee counts reduces the sample to 139 firms. Requiring each firm to appear in both the pre-period (2021–2023) and the post-period (2024) removes a further seven firms, giving a final estimation sample of 132 firms and 487 firm-year observations, of which 69 fall below and 63 above the 500-employee threshold. Table 2 reports the step-by-step attrition.

The panel is unbalanced: 98 of the 132 firms are observed in all four years. We examine the implications in Section 4.3 through a balance comparison of 2021-entry firms with later entrants and present a fully balanced-panel robustness check in Section 6.5. Of the 208 firms with successful text extraction, 98 also appear in the universe sample. The remaining 110 firms are absent from the universe, primarily because they fall below the 500-employee cutoff that effectively defines the lower bound of LSEG’s ESG-covered universe.

Treatment assignment and controls. Employee count, the variable determining treatment assignment, is time-invariant in our analysis. LSEG reports a single employee figure per firm (the most recent available fiscal-year value), which we apply across all years of the panel. While this is a simplification, it is standard in threshold-based DiD designs; using time-varying assignment would introduce endogeneity concerns if firms strategically adjusted employee counts in response to the regulation. We use log market capitalisation

and log bid-ask spread as time-varying firm-level controls. The bid-ask spread is a standard proxy for market liquidity and information asymmetry (Amihud, 2002); including it addresses the concern that liquidity differences between treated and control firms could confound the estimated treatment effect.

Some firms have missing ESG pillar scores for individual years. We do not fill missing values with zero, since a missing LSEG score indicates that LSEG did not assess the firm in that year, not that the firm scored zero. Instead we apply within-firm linear interpolation for gaps of at most one year. In the final estimation sample, no surviving observation relies on interpolated walk scores; main results are based entirely on directly observed LSEG pillar scores.

3.3 *Talk measure*

We measure sustainability disclosure intensity using a text-based talk vector constructed from firm annual reports via a keyword dictionary. Throughout the paper, talk refers to the volume and vocabulary of sustainability discussion in a firm’s annual report, not to the substantive content, credibility, or specificity of what the firm says it is doing. The dictionary builds on the ESG keyword library of Andrikogiannopoulou et al. (2022), extended with CSRD-specific terms related to ESRS reporting standards, climate transition planning, and regulatory language. The complete keyword list and weighting scheme are provided in Appendix A.

The dictionary is positive-only by design: it captures the extent to which firms make sustainability claims, commitments, and disclosures, but does not penalise firms for discussing negative events or acknowledging shortcomings. In a greenwashing study, the talk vector should measure how much the firm claims to be doing about sustainability, not how optimistically it discusses the topic. Including negative keywords would conflate disclosure quality with disclosure quantity, which is not the object of our measurement.

The dictionary is structured across three dimensions (Environmental, Social, and Governance), enabling both pillar-specific and composite analysis. Within each dimension, keywords carry one of two weights: weight 2 for strong, explicit commitments or hard targets, and weight 1 for softer disclosure, strategy, or policy language. The weight reflects the type of vocabulary used, not the substantive strength of the firm’s underlying commitment; the dictionary cannot distinguish between an aspirational mention of “net zero” and a detailed decarbonisation roadmap. To avoid counting negated claims, we implement a window-based negation check that scores zero on keyword matches falling within five tokens of negation words. Full details are provided in Appendix B.

3.4 *Walk measure*

The walk measure requires an independent, third-party assessment of sustainability performance. We use LSEG’s ESG pillar scores. LSEG collects over 870 company-level ESG measures from public sources and uses a subset of 186 measures, selected for cross-firm comparability and industry materiality, to compute scores across ten thematic categories. These roll up into three pillar scores (Environmental, Social, Governance); each is a percentile rank from 0 to 100, benchmarked against industry peers for the Environmental and Social pillars and against country of incorporation for Governance.

We use the base pillar scores rather than the controversy-adjusted ESGC scores because

the ESGC discount uses size-dependent weights that would mechanically interact with our employee-threshold identification strategy. Also, we prefer aggregated pillar scores to raw firm-reported emissions, given the measurement concerns documented by [Aswani et al. \(2024\)](#). One limitation is that some inputs to the pillar scores are Boolean indicators that overlap with what our talk vector captures, creating talk-in-walk contamination. Two points mitigate the concern. The scores also incorporate quantitative indicators (actual CO₂ emissions, water and energy use, environmental R&D) that reflect operational outcomes, and any contamination biases against finding a gap, since it pushes talk and walk in the same direction. Our gap estimates are therefore a conservative lower bound on the true misalignment.

We define composite walk as the arithmetic mean of the three pillar scores, with equal weights to maintain symmetry with our equally-weighted composite talk. We also run all specifications separately on the Environmental Pillar Score, where we expect the strongest treatment effect given ESRS’s heavy emphasis on environmental and climate-related reporting. The Environmental Pillar aggregates emissions, resource use, and innovation category scores, percentile-ranked within each firm’s industry group; further detail on category weights and construction is in [Appendix B](#).

3.5 The disclosure–performance gap

The gap measure captures the misalignment between disclosure intensity and assessed performance. Constructing it requires placing talk and walk on a common scale. The raw measures are fundamentally different: talk is keyword frequency per 1,000 words, while LSEG scores are percentile ranks. We standardise both as z -scores using means and standard deviations from the pre-period (2021–2023) only:

$$z_{i,t} = \frac{x_{i,t} - \bar{x}_{\text{pre}}}{\sigma_{\text{pre}}}, \quad (1)$$

where $\bar{x}_{\text{pre}}$ and σ_{pre} are the cross-sectional mean and standard deviation across all firm-year observations in 2021–2023. Using the pre-period as reference ensures post-period values reflect deviations from the pre-CSR baseline rather than being mechanically re-centred by the treatment itself. We define the gap for each ESG dimension $d \in \{E, S, G\}$ as

$$\text{gap}_{d,i,t} = z \text{ talk}_{d,i,t} - z \text{ walk}_{d,i,t}. \quad (2)$$

A value of zero means that talk and walk are aligned relative to the pre-period distribution; a positive value means the firm talks more than it walks; a negative value means performance exceeds rhetoric. The composite gap is constructed analogously from composite talk and walk.

The gap is symmetric in construction but not in economic content: an increase can arise either because talk rose faster than walk, or because walk declined while talk held constant. We decompose the sources of gap changes in [Section 4.4](#). For the universe descriptive sample the gap is computed identically but using the structured LSEG talk proxy paired with the Environmental Pillar Score.

3.6 Limitations of the measurement approach

LSEG’s ESG coverage, while extensive, is not universal; the unbalanced panel reflects this and Section 6.5 tests robustness to a strict balanced-panel restriction. Moreover, industry classification is not available in our LSEG extraction for the DiD sample, which prevents us from including industry fixed effects; firm fixed effects in our TWFE specification absorb time-invariant industry membership but do not absorb sector-specific time trends. Finally, pillar scores are percentile ranks within industry peer groups, so our finding that walk does not move in response to CSRD is a statement about relative rank within industry, not about absolute environmental performance.

4. Empirical Strategy and Main Results

4.1 Difference-in-differences design

We estimate the effect of CSRD on the talk–walk gap using a difference-in-differences design that exploits the directive’s employee-based scope threshold. CSRD Wave 1 applies to publicly listed European firms with more than 500 employees, so firms above this threshold constitute the treatment group and firms below it the control group. The identifying assumption is that, absent CSRD, the talk–walk gap would have evolved similarly for both groups; we assess this assumption formally in Section 4.3. CSRD was adopted in December 2022, with Wave 1 requiring ESRS-based reporting for financial year 2024 onward. Our pre-period therefore spans 2021–2023 and the post-period is 2024.

Our main estimating equation is a two-way fixed effects model:

$$\text{gap}_{i,t} = \alpha_i + \gamma_t + \beta \cdot (\text{Treated}_i \times \text{Post}_t) + X'_{i,t}\Gamma + \varepsilon_{i,t}, \quad (3)$$

where α_i and γ_t are firm and year fixed effects, $\text{Treated}_i = \mathbf{1}\{\text{employees}_i > 500\}$, $\text{Post}_t = \mathbf{1}\{t \geq 2024\}$, and $X_{i,t}$ includes log market capitalisation and log bid-ask spread as time-varying controls. We cluster standard errors at the firm level. The coefficient of interest is β , capturing the differential change in the talk–walk gap for treated firms relative to controls after CSRD. A positive β indicates that the gap widened more for treated firms; their sustainability talk increased relative to their walk by more than for comparable untreated firms.

We estimate this specification on two samples. The main sample uses a 125–1,000-employee bandwidth (132 firms, 487 observations); the narrower robustness sample uses a 250–750-employee bandwidth that improves comparability between treated and control firms at the cost of statistical power (82 firms, 303 observations). For each bandwidth we estimate the model using both the composite gap and the environmental gap as dependent variables.

Threats to identification. We identify three main threats to causal identification. The first is anticipation. CSRD was adopted in December 2022, giving firms over a year to prepare before the FY2024 reporting requirement bound. If treated firms began adjusting their sustainability language before formal implementation, the estimated treatment effect would be attenuated. The dynamic DiD in Section 4.3 and the 2023 placebo in Section 6.6 bear directly on this; we interpret any resulting bias as making our main estimate a

conservative lower bound on the true first-year effect.

The second is threshold manipulation. If firms strategically adjust their employee count to remain below the 500-employee cutoff, treatment assignment becomes endogenous. Our use of time-invariant employee data mitigates this concern, as does the fact that the threshold was defined by pre-existing NFRD scope rules rather than introduced de novo by CSRD.

The third is spillovers (SUTVA). If control firms voluntarily adopted ESRS-style reporting in response to competitive pressure or in anticipation of future waves, the treatment-control contrast would be compressed. Any such spillover biases β toward zero, making our estimates conservative rather than inflated.

The 500-employee rule is a sharp threshold, but we do not implement a regression-discontinuity design. The 125–1,000-employee bandwidth is wide by RD standards, the running variable is time-invariant, and we have only 132 firms. Our design is a threshold-based DiD with a cross-sectional treatment indicator, not a local polynomial around a discontinuity. The robustness of the main coefficient to the narrower 250–750 bandwidth (Section 6.1) speaks to the same concern that RD would address.

4.2 Main results

Summary statistics for the variables used in the DiD analysis are reported in Table 3. Table 4 reports the main estimates. In the main 125–1,000 specification, the composite talk-walk gap widens by 0.648 standard deviations for treated firms relative to controls after CSRD ($p = 0.026$), and the environmental gap widens by 0.841 standard deviations ($p = 0.016$). The larger environmental effect is consistent with ESRS’s heavy emphasis on environmental and climate-related reporting, which requires extensive disclosure on topics such as climate change mitigation, pollution, and biodiversity.

The 250–750 narrow-bandwidth specification yields coefficients of the same sign and comparable magnitude (+0.433 for the composite gap, +0.555 for the environmental gap) but neither reaches conventional significance. The directional consistency, combined with a 38% reduction in sample size, suggests that the loss of significance reflects imprecision rather than a qualitatively different finding.

Among the controls, log market capitalisation enters negatively and is highly significant in the composite specifications, indicating that larger firms tend to have smaller talk-walk gaps. This is consistent with the universe-level cross-sectional evidence in Section 5 and with the intuition that larger firms face greater scrutiny and have more resources for substantive environmental programmes. Log bid-ask spread is negative but insignificant; adding the spread control slightly strengthens rather than weakens the main results, addressing concerns raised in prior literature about the role of market microstructure in ESG research.

A positive β does not imply that CSRD increased greenwashing in a normative sense. As the decomposition in Section 4.4 shows, the widening is driven by a sharp increase in treated firms’ sustainability language rather than by a deterioration in their environmental performance. Whether this compliance-driven talk surge represents genuine greenwashing or a transitional adjustment depends on whether walk catches up in subsequent reporting years, a question our single post-period cannot answer.

4.3 Dynamic DiD and pre-trends

To assess the timing of the treatment effect and test the parallel-trends assumption, we replace the single Treated $\times$ Post interaction with year-specific interactions, using 2023 as the reference year:

$$\text{gap}_{i,t} = \alpha_i + \gamma_t + \sum_{\tau \neq 2023} \beta_\tau \cdot (\text{Treated}_i \times \mathbf{1}\{t = \tau\}) + X'_{i,t}\Gamma + \varepsilon_{i,t}, \quad (4)$$

where $\tau \in \{2021, 2022, 2024\}$. Each β_τ captures the differential gap between treated and control firms in year τ relative to 2023. Under parallel trends, $\beta_{2021} = \beta_{2022} = 0$; a significant β_{2024} alongside insignificant pre-period coefficients supports a causal interpretation.

The 2021 coefficients are statistically insignificant for both the composite gap ($\beta = -0.218$, $p = 0.358$) and the environmental gap ($\beta = +0.245$, $p = 0.280$), and the 2022 composite coefficient is similarly insignificant ($\beta = -0.275$, $p = 0.161$). The primary pre-trend test is therefore passed: in the years immediately preceding treatment, treated and control firms show no differential trajectory in the composite talk-walk gap.

The 2022 environmental coefficient is conventionally significant and negative ($\beta = -0.318$, $p = 0.043$), indicating that treated firms' environmental gap was trending modestly downward relative to controls in the year before CSRD took effect. Its sign (a downward drift before a sharp upward jump in 2024) runs in the opposite direction from the post-treatment effect, so if anything it makes the post-CSRD reversal more conservative: the environmental gap was narrowing before the directive bound and then widened sharply once it did.

The 2024 coefficients confirm the static DiD finding. The composite gap rises by $\beta = +0.492$ ($p = 0.075$) and the environmental gap by $\beta = +0.662$ ($p = 0.043$). The dynamic-DiD coefficients are smaller than their static counterparts (0.648 and 0.841 respectively) because the two specifications test different contrasts: the static DiD compares the post-period to the average of all three pre-period years, while the dynamic specification compares 2024 to the 2023 reference year only. With our unbalanced panel and a single post-period observation, the two need not coincide exactly. The sharp 2024 jump in talk, alongside the lack of movement in walk, is visible in Figure 3. A cluster-robust Wald test of the joint null $\beta_{2021} = \beta_{2022} = 0$ fails to reject for both specifications: $W = 1.99$ ($p = 0.370$) for the composite gap and $W = 4.16$ ($p = 0.125$) for the environmental gap. Parallel trends are formally supported in both cases.

4.4 Talk-walk decomposition

Is the widening gap driven by treated firms increasing their language, by a relative decline in their performance, or by both? To decompose, we estimate the main TWFE specification separately with standardised talk and standardised walk as dependent variables. By construction, $\beta_{\text{gap}} \approx \beta_{\text{talk}} - \beta_{\text{walk}}$. Table 5 reports the treatment coefficients.

For the composite dimension, treated firms' talk rose by 0.694 standard deviations relative to controls after CSRD ($p = 0.012$), while their walk showed no significant differential change ($+0.046$, $p = 0.582$). The environmental dimension shows the same pattern: talk rose by 0.829 SD ($p = 0.013$), while walk was essentially unchanged (-0.012 , $p = 0.892$). The widening is driven entirely by a surge in sustainability language; environmental per-

formance did not move. The dynamic version of this decomposition (Figure 3) shows that talk coefficients are modestly negative but statistically insignificant in 2021 and 2022 before jumping sharply in 2024, while walk coefficients are indistinguishable from zero throughout.

Two interpretations are consistent with $\beta_{\text{walk}} \approx 0$ and $\beta_{\text{talk}} > 0$. Under compliance-driven disclosure expansion, ESRS requires extensive reporting on environmental policies, emission targets, climate risk assessments, and transition plans, topics that overlap directly with the vocabulary our dictionary counts. Firms newly subject to CSRD must articulate these topics regardless of whether their underlying environmental investments have changed, mechanically expanding the textual footprint of sustainability. Under deliberate overclaiming, firms exploit the expanded ESRS framework to make claims beyond what compliance requires, using mandated reporting as cover for reputational benefit. Distinguishing the two requires either a claims-level text measure that isolates substantive commitments from mandated disclosure, or a longer post-period during which performance has had time to respond. Our measure cannot do the first and our data do not yet allow the second; the heterogeneity evidence in Section 4.5 provides indirect support for the compliance interpretation.

4.5 Heterogeneity

We re-estimate the main TWFE specification on subsamples defined by binary partitions. For a partition into groups A and B , we estimate $\text{gap}_{i,t} = \alpha_i + \gamma_t + \beta_g \cdot (\text{Treated}_i \times \text{Post}_t) + X'_{i,t}\Gamma + \varepsilon_{i,t}$ separately for $g \in \{A, B\}$. We implement three splits: market capitalisation (above versus below the sample median), country (Nordic versus non-Nordic), and pre-period gap (above versus below the median pre-period composite gap). We split on market capitalisation rather than employees because splitting on employees would be mechanically circular with treatment assignment. Table 6 reports the results.

Firms with above-median market capitalisation show larger coefficients than those below the median, consistent with larger firms facing greater analyst and investor scrutiny and more pressure to demonstrate ESRS compliance. Smaller subsample sizes within each split reduce statistical power relative to the full-sample specification. Nordic firms exhibit positive but statistically insignificant coefficients, while non-Nordic firms show large and highly significant coefficients. Nordic countries, particularly Sweden and Finland, have historically maintained stronger sustainability reporting traditions, with many firms already voluntarily reporting under frameworks such as GRI or TCFD before CSRD. For these firms the marginal increase in sustainability language required by ESRS compliance was smaller. Non-Nordic firms (predominantly headquartered in Germany, Italy, and Spain) had more ground to cover, producing a larger talk surge.

If the gap widening reflected deliberate overclaiming, the response would concentrate among firms with high pre-period gaps. We find the opposite. The effect is largest for firms with low pre-period gaps, whose talk and walk were previously well aligned and who therefore had the most disclosure ground to cover under ESRS. Firms with above-median pre-period gaps show coefficients of the same sign but substantially smaller magnitude and no statistical significance. This is the signature of compliance-driven disclosure expansion, not strategic greenwashing, and it is consistent with the country split: non-Nordic firms, who had less developed voluntary reporting before CSRD, show the largest effect.

Taken together, the heterogeneity results reveal that the average treatment effect masks meaningful variation across observable firm characteristics. The CSRD-induced widening is concentrated among non-Nordic firms and firms with low pre-period gaps. The pattern is consistent with a compliance mechanism in which the talk surge is largest for firms facing the greatest adjustment burden under the new reporting requirements.

5. Cross-Sectional Universe Evidence

To complement the main DiD analysis, we examine descriptive and cross-sectional evidence from the universe sample of 981 European listed firms. Because extracting and processing annual reports for nearly 1,000 firms is infeasible, the talk measure in this sample is the structured LSEG environmental proxy described in Section 3.5, with full construction in Appendix B. Because the universe sample contains essentially no firms below the 500-employee threshold, the treated-control comparison underlying the DiD cannot be replicated; we use this sample for descriptive and correlational analysis only.

The baseline cross-sectional specification is

$$\begin{aligned} \text{gap}_{E,i,t} = & \delta_0 + \delta_1 \log(\text{employees}_i) + \delta_2 \log(\text{market cap}_{i,t}) \\ & + \delta_3 \log(\text{spread}_{i,t}) + \mu_c + \lambda_t + u_{i,t}, \end{aligned} \quad (5)$$

where μ_c are country fixed effects and λ_t are year fixed effects. For single-year cross-sectional specifications we use HC1 heteroskedasticity-robust standard errors; for the panel specification we cluster at the firm level.

5.1 Descriptive trends

Both talk and walk exhibit a clear upward trend over the observation period. The mean environmental talk proxy increased from 53.6 in 2021 to 57.2 in 2024, and the mean Environmental Pillar Score (walk) rose from 52.5 to 56.0 over the same period (Figure 4). Talk consistently exceeds walk but the two move broadly in parallel. The resulting gap hovers near zero across all four years, with a slight positive spike in 2022 and a return to baseline in 2023–2024 (Figure 5). This pattern contrasts with the DiD estimation sample, where the gap widens markedly in 2024 for treated firms. The contrast is expected, since the universe sample is dominated by large firms (over 5,000 employees) for which the gap tends to be negative.

5.2 Size–gap relationship

Table 7 reports the cross-sectional regression results. Larger firms have significantly smaller environmental gaps across all specifications: the coefficient on $\log(\text{employees})$ ranges from -0.144 in the univariate specification to -0.046 in the most saturated panel specification, and is statistically significant in every case. Log market capitalisation is consistently negative and significant, reinforcing the size–gap relationship through an alternative size measure. The size-group descriptive statistics confirm this pattern: firms above 5,000 employees exhibit negative average gaps (walk exceeds talk), while firms in the 500–1,000 employee range show positive gaps. Larger firms tend to have more developed environmental programmes and face greater scrutiny from analysts, investors, and

regulators, so their performance is more likely to exceed their stated commitments rather than the reverse.

5.3 Liquidity

$\log(\text{bid-ask spread})$ is significantly associated with the gap in bivariate regressions but the association drops to near zero and becomes insignificant once country fixed effects are included. This suggests that the unconditional spread–gap correlation is driven by cross-country differences in market liquidity, with Nordic markets typically more liquid and showing different average gap levels than Southern European markets, rather than by a firm-level relationship between liquidity and greenwashing. The finding is directly relevant to the main DiD: it provides additional reassurance that the bid-ask spread control absorbs country-level liquidity variation without confounding the estimated treatment effect.

5.4 Returns

We regress annual total returns on the environmental gap, controlling for log market capitalisation, log employees, and country fixed effects. The gap coefficient is $+0.009$ ($p = 0.33$), small, positive, and statistically insignificant. The market does not appear to differentially price high-gap firms at the annual frequency. While this is a null result it is worth reporting: the talk–walk misalignment captured by our greenwashing proxy, despite varying systematically with firm characteristics, is not yet reflected in equity returns. One possible explanation is that the market has not had sufficient time to incorporate the first year of CSRD-mandated disclosures into pricing. This is consistent with [Bolton and Kacperczyk \(2021\)](#), who find that carbon risk is priced in equity markets but note that the pricing of sustainability-related information evolves as disclosure quality improves. We treat this test as descriptive.

5.5 Summary

The universe evidence provides useful context. The environmental greenwashing proxy varies systematically with firm size and geography, with larger firms and firms in countries with stronger reporting traditions exhibiting smaller or negative gaps. The secular upward trend in both talk and walk across the full European market suggests that the aggregate improvement in sustainability disclosure is a market-wide phenomenon, which the DiD year fixed effects absorb. Together, these patterns are consistent with the DiD results and support our interpretation that CSRD’s impact on the gap operates through the talk channel rather than reflecting broader shifts in the European sustainability landscape.

6. Robustness

This section assesses whether the main DiD results are sensitive to the specification choices made throughout the analysis. We examine six dimensions: bandwidth choice, inclusion of controls, outlier influence, the compositionally distinct 2021 subsample, restriction to a balanced panel, and placebo timing. Section 6.7 summarises coefficient stability across all six. Detailed estimates are reported in Tables 8 and 9.

6.1 Bandwidth sensitivity

The main specification uses a 125–1,000-employee bandwidth. We also estimate the model on a narrower 250–750-employee bandwidth that improves comparability between treated and control firms at the cost of sample size. For the narrow sample, the composite coefficient remains positive ($\beta = +0.433$) but does not reach statistical significance ($p = 0.331$), and the environmental coefficient follows the same pattern. Coefficient signs are identical to the main specification and magnitudes are around 30% smaller but in the same range. We attribute the reduced precision to the smaller sample (82 firms and 303 observations) rather than to a substantively different effect. Directional consistency across bandwidths suggests that the main finding is not an artefact of including firms far from the threshold.

6.2 Sensitivity to controls

We re-estimate the TWFE specification under three configurations: no controls; log market capitalisation only; and the full specification with both log market capitalisation and log bid-ask spread. Coefficients are highly stable. The unconditional specification yields a composite coefficient of $+0.607$ ($p = 0.041$) and an environmental coefficient of $+0.799$ ($p = 0.025$). Adding log market capitalisation alone produces nearly identical estimates. The full specification slightly strengthens the result. Across all three configurations, coefficients vary by less than 0.05 and significance is preserved throughout. The main effect is not sensitive to the choice of controls; neither liquidity nor firm size confounds the treatment effect.

6.3 Outlier robustness

To assess whether extreme observations drive the result, we winsorise the gap variables at the 5th and 95th percentiles and re-estimate the main specification. The composite gap remains positive and significant ($\beta = +0.640$, $p = 0.002$), as does the environmental gap ($\beta = +0.775$, $p = 0.001$). Significance increases relative to the unwinsorised specification, indicating that the main result is not driven by outliers; removing extreme values reduces residual variance and produces a more precise estimate of the average treatment effect.

6.4 Excluding the 2021 subsample

2021 was the first full post-COVID reporting year, during which corporate disclosure patterns may have been atypical. As a sensitivity check, we re-estimate the main DiD restricting the sample to 2022–2024 only. The results yield coefficients of the same sign and comparable magnitude with statistical significance preserved (composite: $\beta = +0.617$, $p = 0.041$; environmental: $\beta = +0.808$, $p = 0.028$). These estimates are within 0.05 of the main specification, confirming that the main result is not driven by post-COVID-specific dynamics.

6.5 Balanced panel

The main panel is unbalanced: while 132 firms enter the estimation sample, only 98 are observed in all four years. The remaining 34 firms appear in three or fewer years, primarily because their earliest annual reports were unavailable. To investigate whether this compositional variation affects the results, we re-estimate the main DiD using only the 98

firms observed across all of 2021–2024. Coefficient signs and approximate magnitudes are preserved, but significance weakens (composite: $\beta = +0.470$, $p = 0.160$; environmental: $\beta = +0.648$, $p = 0.095$). The composite gap loses significance at conventional thresholds, while the environmental gap remains marginally significant at the 10% level.

This is the most demanding robustness check we conduct. The drop in significance reflects two compounding factors: a 20% reduction in panel size (from 488 to 391 firm-year observations) and the loss of firms whose entry timing carried informative variation about the treatment effect. We retain the unbalanced panel as our main specification because it preserves meaningful variation that a strict balanced restriction discards, and because the central finding holds across the other five robustness dimensions. We acknowledge that the balanced-panel result introduces some uncertainty about the magnitude, though not the direction, of the treatment effect.

6.6 Placebo tests

As a further check that the estimated treatment effect is specific to the actual CSRD implementation date, we conduct placebo tests on the pre-period. We restrict the sample to the pre-period (2021–2023), exclude all post-CSRD observations, and re-estimate the DiD model as if the directive had been implemented for 2022 and 2023 respectively. Under the null, these regressions should yield insignificant coefficients. Table 9 reports the results.

Neither placebo year produces a statistically significant treatment effect. With 2022 as the hypothetical policy year, the composite and environmental gap coefficients are +0.078 and +0.104 respectively, both small in magnitude and far from significance. With 2023 as the hypothetical policy year, the coefficients are somewhat larger (+0.219 for composite and +0.249 for environmental) but remain insignificant. The 2023 environmental coefficient ($p = 0.171$) approaches marginal significance and could reflect early anticipation effects, with firms aware of upcoming CSRD requirements beginning to adjust their reporting language in 2023. Any such anticipation biases our main 2024 coefficient downward by shifting part of the talk adjustment into the pre-period, and the dynamic DiD (with 2023 as the reference year) absorbs any 2023 adjustment into the baseline. This makes the main estimate a conservative lower bound on the true first-year treatment effect. Since neither placebo coefficient meets conventional significance thresholds, the placebo tests support the interpretation that the statistically significant effect we observe is tied to the actual policy implementation in 2024.

6.7 Summary

The main results are robust across all six dimensions tested. Coefficient signs are preserved in every specification, and magnitudes vary by less than 0.2 standard deviations across most checks. The two specifications where statistical significance weakens (narrow bandwidth and balanced panel) reflect reduced statistical power from substantially smaller samples rather than qualitatively different findings, and in both cases the coefficient direction and approximate magnitude are preserved. Taken together, these results support the conclusion that the CSRD-induced widening of the talk–walk gap reflects a genuine treatment effect rather than an artefact of specific modelling choices.

7. Discussion

7.1 *Interpreting the positive coefficient*

The positive β is consistent with the compliance channel dominating in the first reporting year. It does not imply that CSRD failed or that greenwashing increased in a normative sense. The decomposition shows that walk did not deteriorate; firms subject to CSRD are not performing worse environmentally, they are disclosing more extensively. The question is therefore not whether CSRD caused harm, but whether the talk surge it triggered reflects genuine regulatory compliance (firms fulfilling their ESRS reporting obligations) or strategic overclaiming (firms inflating sustainability language beyond what compliance requires). Our data cannot distinguish these within the current observation window.

The compliance interpretation is reinforced by the heterogeneity results. The effect is concentrated among firms that had the most adjustment to make: firms with low pre-period gaps and firms headquartered outside Nordic countries. This pattern is more consistent with adaptation to new regulatory requirements than with strategic behaviour, since deliberate greenwashers would presumably be found among firms with high pre-period gaps rather than among firms whose talk and walk were previously aligned.

Our findings speak to the framework of [Inderst and Opp \(2025\)](#), who argue that a mandatory taxonomy is necessary to discipline greenwashing. Their analysis implicitly assumes that forced disclosure constrains cheap talk; our evidence suggests a nuanced sequencing. In the short run, forced disclosure first increases the volume of sustainability language as firms adapt to a new reporting framework, before the disciplining mechanism takes hold in subsequent reporting cycles. Whether this disciplining effect eventually materialises, narrowing the gap as walk catches up to talk, is the central question open to future research.

7.2 *Short-run versus long-run effects*

The most consequential limitation of our analysis is that we observe only one post-treatment year. Our estimates are therefore inherently short-run. Under the compliance interpretation, the gap should narrow in subsequent years as two forces converge: treated firms' walk catches up through actual environmental investments, and the initial language adjustment plateaus, since firms can only adopt ESRS terminology once and the marginal talk increase should diminish in later reporting cycles. Under the deliberate overclaiming interpretation, the gap would instead persist or widen further as firms continue to exploit the expanded disclosure framework for reputational benefit without corresponding action. Distinguishing between these trajectories requires extending the analysis to FY2025 and beyond, a natural and important direction for future research.

7.3 *External validity*

Our sample is restricted to European listed firms with 125–1,000 employees, a narrow size band around the CSRD threshold. Results may not generalise to very large firms, which were already subject to NFRD reporting and had established sustainability infrastructure, or to non-listed firms, which face different disclosure incentives and market pressures. The universe cross-sectional evidence in [Section 5](#) provides partial reassurance: the size-gap relationship at the threshold is consistent with broader patterns across the European

market, where larger firms systematically show smaller or negative gaps.

Geographic concentration is also relevant. Around 27% of our DiD sample consists of Swedish firms, with Germany and Finland as the next-largest country groups. The country heterogeneity analysis showed that the effect is concentrated in non-Nordic firms, but coverage within that group is dominated by a few countries. Results may differ for Southern or Eastern European firms with less developed capital markets and different regulatory traditions. Sector-level heterogeneity also remains untested: the absence of industry classification in our LSEG extraction prevents us from examining whether the treatment effect differs across high- versus low-emission industries, an important dimension given that ESRS imposes particularly detailed requirements on climate-intensive sectors.

7.4 Measurement limitations

Several measurement choices could influence interpretation. The talk measure, while grounded in [Andrikogiannopoulou et al. \(2022\)](#) and extended with CSRD-specific terminology, is a simplified representation of corporate sustainability language. It captures keyword frequency but not semantic content, rhetorical tone, or the specificity of claims. A firm that mentions “net zero” once in a vague aspirational statement receives the same weight-2 score as one providing a detailed, science-aligned decarbonisation roadmap. Moreover, the positive-only design of the dictionary may amplify the measured compliance effect, since ESRS compliance generates precisely the kind of affirmative policy and target language the dictionary is designed to capture. More sophisticated natural-language-processing approaches, for instance large language models trained on sustainability disclosure, could capture qualitative distinctions in disclosure quality but are beyond the scope of this thesis.

LSEG’s Environmental Pillar Scores aggregate multiple indicators using proprietary weighting, which introduces opacity into the walk measure. Pillar scores are also relative rankings within industry peer groups: a firm’s walk score reflects its position relative to peers rather than its absolute environmental impact. A firm whose emissions increase in absolute terms may nonetheless see its walk score improve if industry peers deteriorate faster, a subtlety that complicates the interpretation of walk as a pure measure of environmental performance. Finally, the gap construction is a simple linear difference of z -scores, which treats a one-unit increase in talk identically regardless of starting level. In practice, the marginal greenwashing content of an additional unit of talk may differ between a firm already disclosing extensively and one previously silent.

7.5 Summary

Despite these limitations, the core finding (that CSRD’s first year produced a significant, compliance-driven widening of the talk–walk gap) rests on a credible identification strategy, is robust to a range of specification choices, and is supported by an economically coherent mechanism. The results contribute to an empirical literature that, to our knowledge, has not yet tested firm-level effects of CSRD using talk–walk gap measures.

8. Conclusion

Wave 1 of the EU’s Corporate Sustainability Reporting Directive did not narrow the gap between what firms say about sustainability and what third parties say they do. In the directive’s first reporting year, the composite talk–walk gap widened by 0.65 standard deviations for treated firms and the environmental gap by 0.84, driven entirely by a surge in disclosure language with no movement in performance. The effect concentrates among firms with the most disclosure ground to cover, namely non-Nordic firms and firms with initially aligned talk and walk, which is the opposite of what deliberate overclaiming would predict.

We interpret the positive coefficient as compliance-driven expansion of mandated disclosure, not as evidence that CSRD increased greenwashing in a normative sense. Treated firms are disclosing more, not performing worse. Two caveats shape the conclusion. Our talk measure captures disclosure intensity rather than the substantive content of firm claims, so a compliance-driven expansion and a deliberate overclaiming response are observationally equivalent under a dictionary-based text measure. We also observe only one post-treatment year. Whether walk catches up from FY2025 onward, or the gap persists, is a question our data cannot yet answer.

The paper contributes to existing research in two ways. Empirically, it provides the first firm-level test of CSRD’s effect on a disclosure–performance gap, documenting a compliance-driven widening in the first reporting year rather than the narrowing that disciplining theories of mandatory disclosure would predict. Methodologically, it flags a measurement issue that any text-based greenwashing study in regulated settings will need to confront: when a disclosure regime mandates vocabulary, dictionary-based measures cannot cleanly separate substantive claims from compliance. Three directions would sharpen what our design cannot resolve. Extending the post-period to FY2025 and beyond would distinguish the compliance and overclaiming trajectories. Claims-level natural-language extraction (using large language models to isolate substantive commitments and targets from mandated risk disclosures) would resolve the interpretive ambiguity. Broader coverage across Southern and Eastern Europe, and across industries with different climate exposures, would speak to external validity in settings our estimation sample does not represent.

References

- Admati, A. R., and P. Pfleiderer, 2000, Forcing firms to talk: Financial disclosure regulation and externalities, *Review of Financial Studies* 13, 479–519.
- Amihud, Y., 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.
- Andrikogiannopoulou, A., P. Krueger, S. F. Mitali, and F. Papakonstantinou, 2022, Discretionary information in ESG investing: A text analysis of mutual fund prospectuses, Swiss Finance Institute Research Paper.
- Aswani, J., A. Raghunandan, and S. Rajgopal, 2024, Are carbon emissions associated with stock returns?, *Review of Finance* 28, 75–106.
- Bolton, P., and M. Kacperczyk, 2021, Do investors care about carbon risk?, *Journal of Financial Economics* 142, 517–549.
- Duchin, R., J. Gao, and Q. Xu, 2025, Sustainability or greenwashing: Evidence from the asset market for industrial pollution, *Journal of Finance*, forthcoming.
- European Commission, 2022, Corporate Sustainability Reporting Directive (Directive (EU) 2022/2464), *Official Journal of the European Union*.
- Hoberg, G., and G. Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.
- Hong, H., and M. Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.
- Ilhan, E., Z. Sautner, and G. Vilkov, 2021, Carbon tail risk, *Review of Financial Studies* 34, 1540–1571.
- Inderst, R., and M. M. Opp, 2025, Sustainable finance versus environmental policy: Does greenwashing justify a taxonomy for sustainable investments?, *Journal of Financial Economics*, forthcoming.
- Krueger, P., Z. Sautner, and L. T. Starks, 2020, The importance of climate risks for institutional investors, *Review of Financial Studies* 33, 1067–1111.
- Pástor, Ľ., R. F. Stambaugh, and L. A. Taylor, 2021, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.
- Sautner, Z., L. van Lent, G. Vilkov, and R. Zhang, 2023, Firm-level climate change exposure, *Journal of Finance* 78, 1449–1498.

Table 1. Universe sample attrition

Construction of the universe descriptive sample from the initial LSEG equity screen to the final panel of 981 European-domiciled listed firms observed 2021–2024. The Clean subset removes firms with insufficient ESG pillar-score coverage; the core-variable filter drops firms with more than one missing year on the Environmental Pillar Score or the structured talk-proxy components. The minimum employee count in the final sample is 504.

Step	Firms	Notes
Initial equity screen	1,475	European-domiciled listed firms, LSEG coverage
“Clean” ESG subset	996	Non-missing ESG pillar scores across 2021–2024
Core-variable coverage	981	Drop firms with >1 missing year for pillar/talk-proxy variables
Final universe panel	981	3,716 firm-year observations

Table 2. DiD sample attrition

Construction of the estimation sample from the initial LSEG employee-bandwidth screen to the final panel of 132 firms. Talk scores are constructed from firm-provided annual reports (see Section 3.3). Walk data are LSEG ESG pillar scores. The pre-period spans 2021–2023; the post-period is 2024. Of the 132 firms, 63 have more than 500 employees (treated under CSRD Wave 1) and 69 fall below that threshold (control).

Step	Firms	Notes
Initial screen (125–1,000 employees)	232	European-domiciled listed firms in bandwidth
Successful text extraction	208	Drop 24 firms without English-language reports or extraction failure
Merge with LSEG metadata	205	Drop 3 firms with identifier mismatches
Bandwidth re-imposed on LSEG employees	139	Drop firms outside 125–1,000 on LSEG counts
Pre- and post-period requirement	132	Drop 7 firms without both pre- and post-period obs
Final DiD panel	132	488 firm-year obs; 69 control, 63 treated

Table 3. Summary statistics

This table reports summary statistics for the variables used in the difference-in-differences analysis. Sample is the 125–1,000 employee bandwidth, 132 firms, 488 firm-year observations, 2021–2024. Panel A reports the dependent and intermediate variables: pillar-level talk scores (keyword hits per 1,000 words of annual report), pillar-level walk scores (LSEG percentile pillar scores), and the standardised composite and environmental talk–walk gaps. Panel B reports the treatment indicator, post indicator, firm characteristics, and the time-varying controls used in the regressions. “Talk” subscripts E, S, G refer to the Environmental, Social, and Governance dimensions. “Bid-ask spread (%)” is the proportional spread defined in Appendix B.4; log market capitalisation is in billions of euros. The unit of observation is firm-year throughout.

Variable	<i>N</i>	Mean	Std. Dev.	5%	25%	Median	75%	95%
<i>Panel A: Talk, walk, and gap variables</i>								
Talk composite	488	1.674	1.375	0.275	0.676	1.294	2.254	4.568
Talk (E)	488	1.122	1.628	0.000	0.076	0.466	1.570	4.308
Talk (S)	488	0.588	0.648	0.007	0.147	0.347	0.851	1.656
Talk (G)	488	3.313	2.356	0.565	1.681	2.807	4.312	7.820
Walk composite	488	41.078	18.633	10.907	27.233	41.472	55.072	70.895
Walk (E)	488	34.226	23.036	0.000	17.705	30.919	50.639	76.080
Walk (S)	488	46.045	21.597	9.736	29.173	47.219	62.606	80.533
Walk (G)	488	42.961	22.007	8.865	25.528	42.923	61.044	78.142
Composite gap	488	0.211	1.390	−1.913	−0.541	0.151	0.799	2.463
Environmental gap	488	0.329	1.463	−1.730	−0.511	0.223	0.733	2.543
<i>Panel B: Treatment, post, and firm controls</i>								
Treated	488	0.471	0.500	0.000	0.000	0.000	1.000	1.000
Post	488	0.270	0.445	0.000	0.000	0.000	1.000	1.000
Employees	488	483.224	238.562	149.500	270.500	482.250	659.500	918.025
Market cap (bn EUR)	488	0.848	1.409	0.015	0.086	0.244	1.018	2.968
Bid-ask spread (%)	488	1.085	1.744	0.089	0.241	0.505	1.338	3.758
log(market cap)	488	−1.324	1.670	−4.226	−2.459	−1.411	0.018	1.088
log(bid-ask spread)	487	−0.571	1.163	−2.394	−1.416	−0.679	0.293	1.324

Table 4. Main difference-in-differences estimates

This table reports estimates from a two-way fixed-effects DiD model where the dependent variable is the standardised talk-walk gap, constructed as the difference between pre-period z -scored sustainability talk (annual report keyword dictionary per 1,000 words) and z -scored LSEG ESG pillar scores for firm i in year t . Columns (1) and (2) report estimates for the composite gap and the environmental gap respectively, both estimated on the main 125–1,000-employee bandwidth. Columns (3) and (4) report the same specifications on the narrower 250–750-employee bandwidth as a robustness check. The coefficient of interest is β on $\text{Treated}_i \times \text{Post}_t$, where $\text{Treated}_i = \mathbf{1}\{\text{employees}_i > 500\}$ and $\text{Post}_t = \mathbf{1}\{t \geq 2024\}$. The Post indicator is collinear with the 2024 year fixed effect and reported here for transparency; year fixed effects are absorbed in all specifications. Standard errors (in parentheses) are clustered at the firm level. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

	125–1,000 (main)		250–750 (narrow)	
	(1)	(2)	(3)	(4)
	Composite gap	Environmental gap	Composite gap	Environmental gap
Treated $\times$ Post	0.648** (0.291)	0.841** (0.348)	0.433 (0.445)	0.555 (0.522)
Post	0.293** (0.132)	0.273* (0.150)	0.449** (0.220)	0.458* (0.249)
log(market cap)	−0.286*** (0.040)	−0.046 (0.046)	−0.227*** (0.049)	−0.097* (0.054)
log(bid-ask spread)	−0.099 (0.078)	−0.084 (0.093)	−0.149 (0.115)	−0.120 (0.129)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firms	132	132	82	82
Observations	487	487	303	303

Table 5. Talk–walk decomposition

This table reports estimates from the main two-way fixed-effects DiD specification, re-estimated separately with standardised talk and standardised walk as dependent variables. Columns (1) and (2) report results for composite talk and walk; columns (3) and (4) report the same for the environmental dimension. By construction, the gap coefficient decomposes approximately as $\beta_{\text{gap}} \approx \beta_{\text{talk}} - \beta_{\text{walk}}$. The coefficient of interest is β on $\text{Treated}_i \times \text{Post}_t$. All specifications include firm and year fixed effects, log market capitalisation, and log bid-ask spread as time-varying controls. Standard errors (in parentheses) are clustered at the firm level. Sample is the 125–1,000 employee bandwidth, 132 firms, 487 firm-year observations. The Memo row reports the direct DiD coefficient when the standardised talk–walk gap is the dependent variable, for comparison. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

	Composite		Environmental	
	(1)	(2)	(3)	(4)
	Talk (z)	Walk (z)	Talk (z)	Walk (z)
Treated $\times$ Post	0.694** (0.275)	0.046 (0.084)	0.829** (0.332)	−0.012 (0.091)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
log(market cap)	Y	Y	Y	Y
log(bid-ask spread)	Y	Y	Y	Y
Memo: β_{gap}	0.648** (0.291)		0.841** (0.348)	
Firms	132	132	132	132
Observations	487	487	487	487
R^2	0.049	0.003	–	–

Table 6. Heterogeneity

This table reports estimates from the main two-way fixed-effects DiD specification re-estimated on subsamples defined by three binary partitions: above versus below median market capitalisation; Nordic (Sweden, Denmark, Finland, Norway, Iceland) versus non-Nordic; and above versus below median pre-period composite gap. For each partition, columns (1) and (2) report the Treated $\times$ Post coefficient and clustered standard error for the composite talk-walk gap; columns (3) and (4) report the same for the environmental talk-walk gap. All specifications include firm and year fixed effects and time-varying controls (log market capitalisation and log bid-ask spread). Standard errors clustered at the firm level. The narrower estimation samples reduce statistical power relative to the full-sample specification in Table 4. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

	Composite gap		Environmental gap		Firms	Obs
	(1)	(2)	(3)	(4)		
	Coef.	SE	Coef.	SE		
<i>Panel A: Market capitalisation</i>						
Below median	0.491	(0.324)	0.723*	(0.389)	66	245
Above median	0.925*	(0.469)	1.077*	(0.567)	66	242
<i>Panel B: Country</i>						
Nordic	0.411	(0.483)	0.717	(0.583)	59	223
Non-Nordic	0.930***	(0.316)	1.010***	(0.377)	73	264
<i>Panel C: Pre-period gap</i>						
Below median (low pre-gap)	1.103***	(0.391)	1.224**	(0.487)	66	244
Above median (high pre-gap)	0.213	(0.449)	0.479	(0.524)	66	243
Firm FE	Y	Y	Y	Y		
Year FE	Y	Y	Y	Y		
log(market cap)	Y	Y	Y	Y		
log(bid-ask spread)	Y	Y	Y	Y		

Table 7. Cross-sectional regressions on the universe sample

This table reports cross-sectional regressions of the standardised environmental talk-walk gap on firm characteristics. The dependent variable is the gap constructed using the structured LSEG environmental talk proxy (see Section 3.3). All independent variables are in natural logs. Column (1) is univariate on log(employees). Column (2) adds log(market cap). Column (3) further adds log(bid-ask spread). Column (4) adds country fixed effects. Column (5) is the panel specification with country and year fixed effects, using firm-clustered standard errors. Columns (1)–(4) report HC1 heteroskedasticity-robust standard errors; column (5) reports standard errors clustered at the firm level. Standard errors in parentheses. Sample is 981 European-domiciled listed firms, 2021–2024, with the panel specification using 3,716 firm-year observations. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Variable	(1) Size only	(2) + Market cap	(3) + Spread	(4) + Country FE	(5) Panel + FEs
log(employees)	−0.144*** (0.017)	−0.072*** (0.021)	−0.068*** (0.023)	−0.059** (0.024)	−0.046** (0.022)
log(market cap)		−0.085*** (0.016)	−0.118*** (0.019)	−0.099*** (0.019)	−0.060*** (0.017)
log(bid-ask spread)			−0.067*** (0.022)	−0.009 (0.028)	0.009 (0.018)
Year FE	No	No	No	No	Yes
Country FE	No	No	No	Yes	Yes
N	981	981	972	972	3,716
R^2	0.070	0.095	0.109	0.156	0.092
Std. errors	HC1	HC1	HC1	HC1	Firm-clustered

Table 8. Robustness summary

This table reports the Treated $\times$ Post coefficient from a series of alternative specifications, alongside the baseline result from Table 4. Columns (1) and (2) report the composite and environmental gap respectively. All specifications include firm and year fixed effects, and time-varying controls (log market capitalisation and log bid-ask spread) except where indicated. The narrow-bandwidth specification restricts the sample to firms with 250–750 employees; the no-controls and market-cap-only specifications progressively remove right-hand-side variables; the winsorised specification trims the dependent variables at the 5th and 95th percentiles; the excluding-2021 specification drops the first sample year; the balanced-panel specification restricts to firms observed in all four years. Standard errors clustered at the firm level in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

	(1)	(2)	Firms	Obs
	Composite gap	Environmental gap		
Main (Table 4)	0.648** (0.291)	0.841** (0.348)	132	487
Narrow bandwidth (250–750)	0.433 (0.445)	0.555 (0.522)	82	303
No controls	0.607** (0.297)	0.799** (0.354)	132	487
Market cap control only	0.604** (0.296)	0.794** (0.354)	132	487
Winsorised (5th/95th pctile)	0.640*** (0.209)	0.775*** (0.232)	132	487
Excluding 2021 obs	0.617** (0.301)	0.808** (0.364)	132	384
Balanced panel (98 firms)	0.470 (0.333)	0.648* (0.387)	98	391
Firm FE	Y	Y		
Year FE	Y	Y		

Table 9. Placebo tests

This table reports placebo tests in which we re-estimate the main two-way fixed-effects DiD model on the pre-period (2021–2023) with the Post indicator set to 1 in a hypothetical placebo policy year (2022 in Panel A; 2023 in Panel B) and 0 otherwise. Columns (1) and (2) report estimates for the composite gap and the environmental gap respectively. The coefficient of interest is β on $\text{Treated}_i \times \text{Post}_t^{\text{placebo}}$. All specifications include firm and year fixed effects and time-varying controls (log market capitalisation and log bid-ask spread). Standard errors clustered at the firm level in parentheses. Under the null that the treatment effect is specific to the actual CSRD implementation date in 2024, both placebo regressions should yield insignificant coefficients. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

	(1)	(2)
	Composite gap	Environmental gap
<i>Panel A: Placebo policy year = 2022</i>		
Treated $\times$ Post ^{placebo}	0.078 (0.186)	0.104 (0.190)
<i>Panel B: Placebo policy year = 2023</i>		
Treated $\times$ Post ^{placebo}	0.219 (0.219)	0.249 (0.181)
Firm FE	Y	Y
Year FE	Y	Y
log(market cap)	Y	Y
log(bid-ask spread)	Y	Y
Observations	356	356

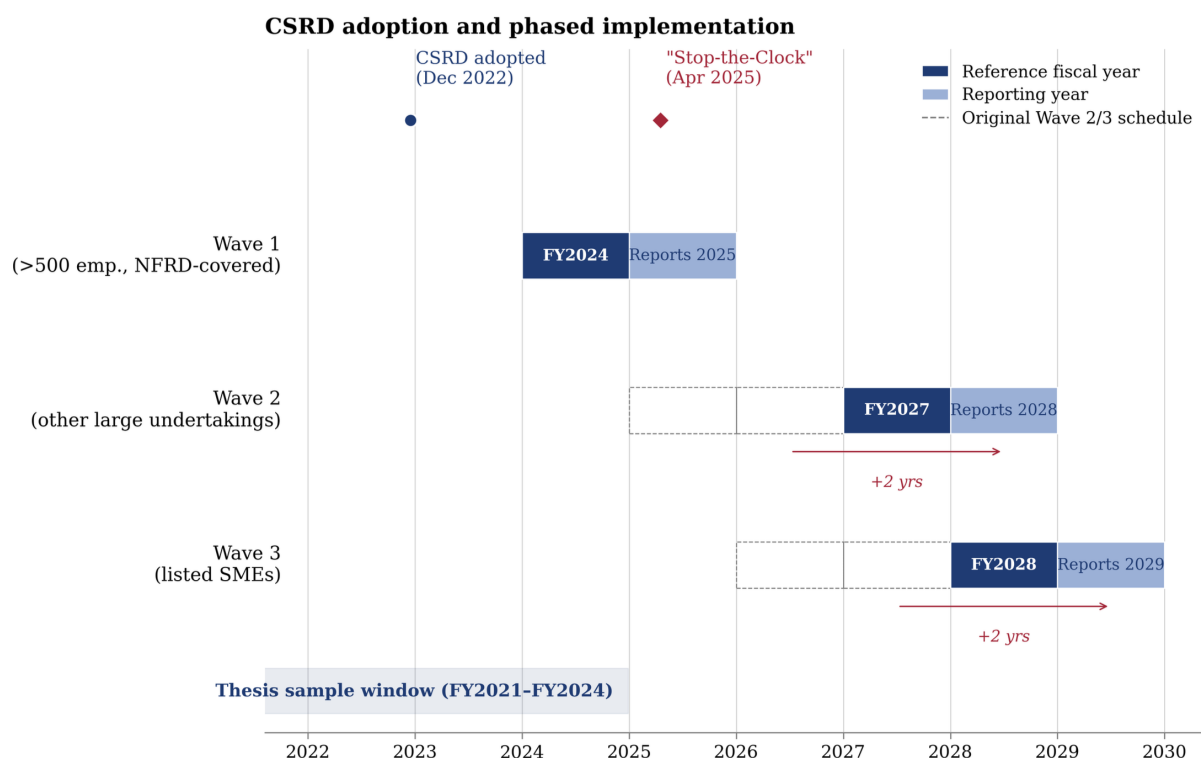


Figure 1. CSRD adoption and phased implementation. The Corporate Sustainability Reporting Directive was adopted by the European Parliament and Council in December 2022 (European Commission, 2022). The European Sustainability Reporting Standards (ESRS) became binding in January 2024. Wave 1 firms (large public-interest entities with more than 500 employees, inheriting NFRD scope) were required to report under ESRS for financial year 2024, with reports published in 2025. Subsequent waves originally intended to extend coverage to smaller listed firms and non-listed undertakings have been postponed by two years under the EU’s April 2025 “Stop-the-Clock” omnibus directive, leaving Wave 1 as the main binding cohort over our analysis window. The shaded band marks the FY2021–FY2024 sample window used in the analysis.

Identification: 500-employee CSRD scope threshold

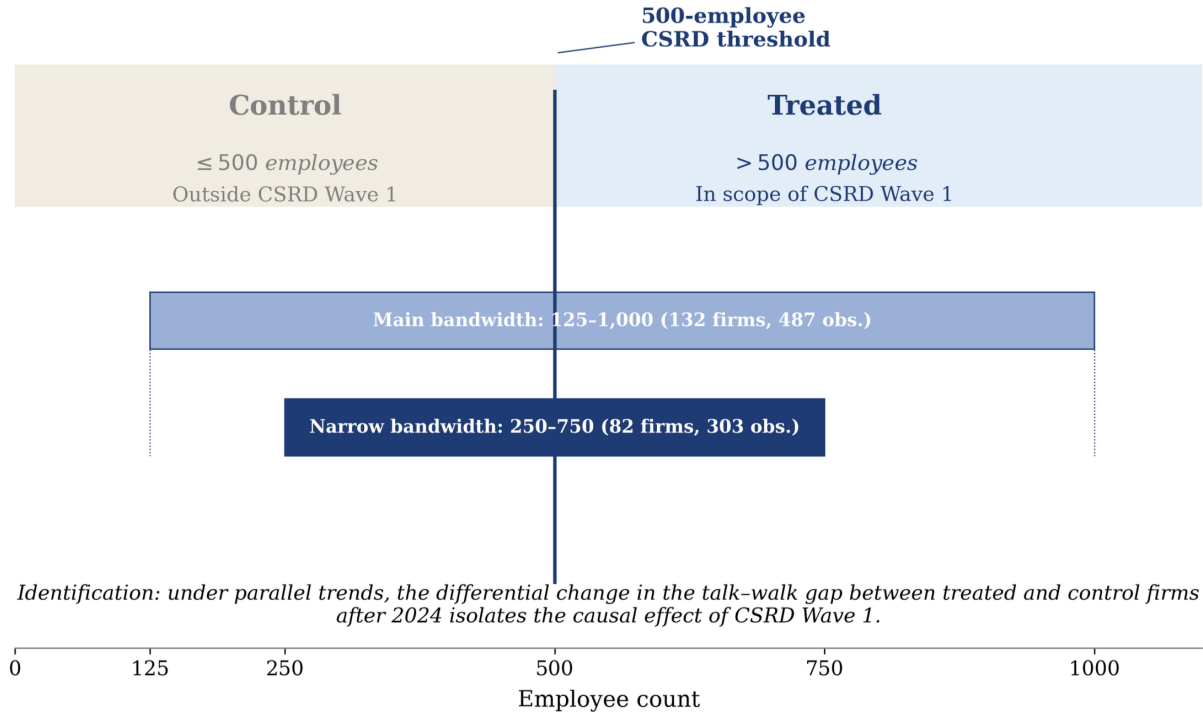


Figure 2. CSRD employee threshold and identification bandwidth. CSRD Wave 1 applies to publicly listed firms with more than 500 employees, the threshold inherited from the earlier Non-Financial Reporting Directive. The estimation sample for the difference-in-differences design is restricted to European-domiciled listed firms with 125 to 1,000 employees, placing treated and control firms in a comparable size range. Treatment assignment is time-invariant: a firm classified as treated retains that classification across all years of the panel. The narrow-bandwidth robustness sample (Section 6.1) tightens the band to 250–750 employees.

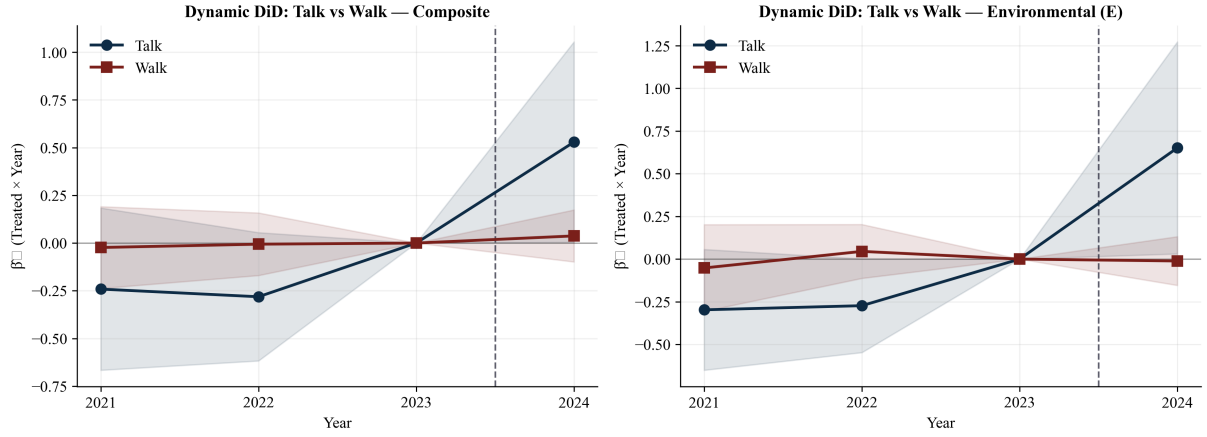


Figure 3. Dynamic DiD: talk versus walk. Year-specific $\text{Treated}_i \times \mathbf{1}\{t = \tau\}$ coefficients with 2023 as the reference year, estimated separately with standardised talk and standardised walk as dependent variables. All specifications include firm and year fixed effects, log market capitalisation, and log bid-ask spread as time-varying controls. Vertical bars indicate 95% confidence intervals based on standard errors clustered at the firm level. The left panel plots the composite dimension; the right panel plots the environmental dimension. In both dimensions, treated firms’ talk shows modestly negative coefficients in 2021 and 2022 before jumping sharply in 2024, while walk coefficients are indistinguishable from zero throughout. The implied gap coefficients (talk minus walk) appear in the body text of Section 4.3.

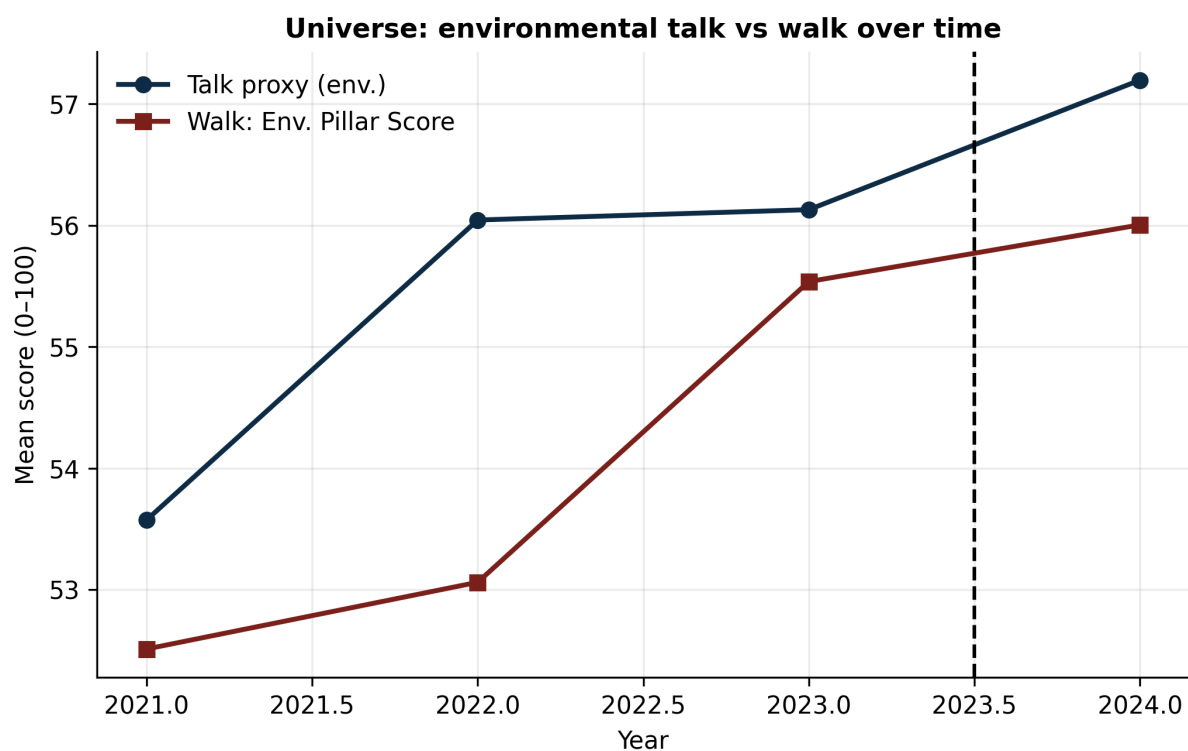


Figure 4. Universe talk and walk over time. Annual means of the structured environmental talk proxy (arithmetic mean of LSEG Policy Emissions, Targets Emissions, and Climate Change Commercial Risks/Opportunities scores) and the LSEG Environmental Pillar Score across 981 European-domiciled listed firms, 2021–2024. Both series are on the 0–100 LSEG percentile scale.

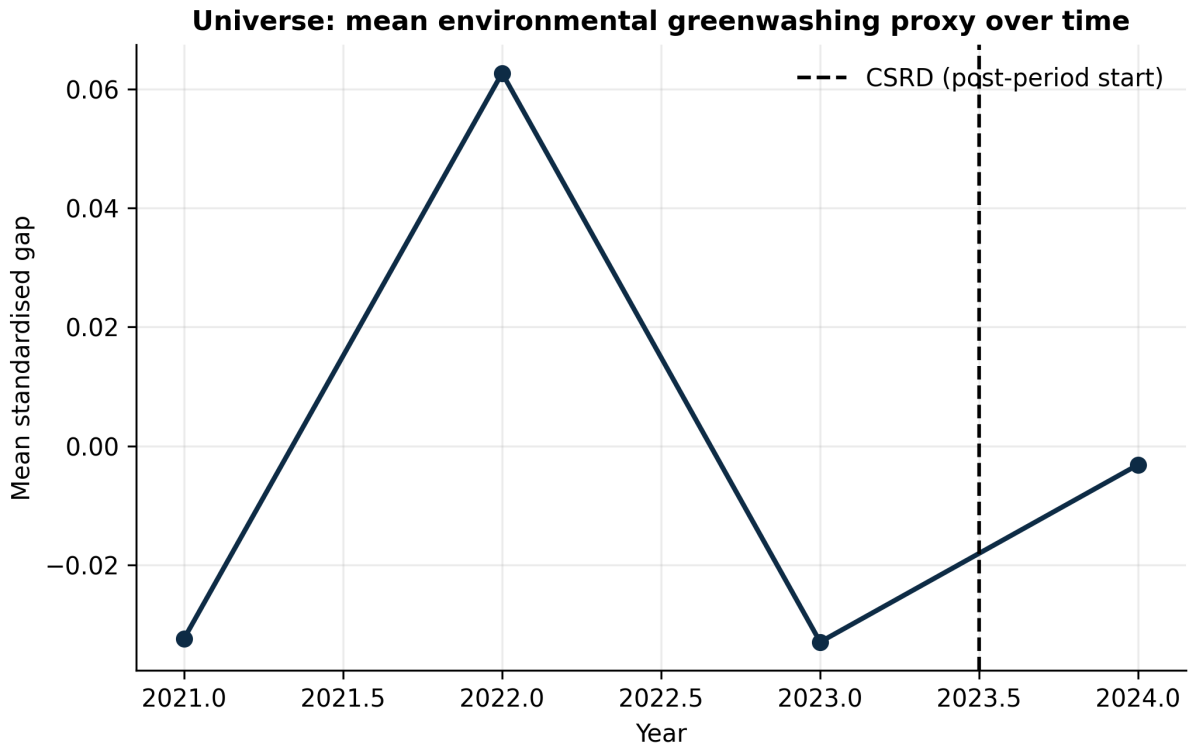


Figure 5. Universe mean greenwashing proxy over time. Annual mean of the universe environmental talk-walk gap, defined as the standardised environmental talk proxy minus the standardised Environmental Pillar Score. Both components are standardised using pre-period (2021–2023) means and standard deviations. A positive value indicates that disclosure intensity exceeds assessed performance.

A. Keyword Dictionary

This appendix documents the full keyword dictionary used to construct the talk measure (Section 3.3). The dictionary builds on the ESG keyword library of [Andrikogiannopoulou et al. \(2022\)](#), extended with CSRD-specific terms related to ESRS reporting standards, climate transition planning, and regulatory language. Keywords are organised across three dimensions (Environmental, Social, Governance) and assigned one of two weights. Weight 2 keywords indicate strong, explicit commitments or hard targets; weight 1 keywords reflect softer disclosure, strategy, or policy language. Keywords appearing in more than one pillar are marked with a dagger (†); a single mention contributes to each pillar it appears in but does not affect the composite disproportionately, since the composite is the arithmetic mean of the three pillars.

The negation logic complements the dictionary. For each keyword match we examine a window of five tokens on either side and check for negation words (“not”, “no”, “never”, “without”, “unable”, “lack”, “failure”). If a negation word is found within this window the keyword match scores zero rather than its dictionary weight. The window does not cross sentence boundaries: a negation word at the end of one sentence cannot nullify a keyword at the beginning of the next. For example, “We have *not* yet set a science-based target” correctly scores zero for “science-based target”, while two adjacent sentences such as “[...] challenges remain. Our science-based target covers Scope 1 and 2 emissions” correctly score the keyword at full weight.

After scoring, raw pillar scores are normalised by dividing by the total word count of the annual report and multiplying by 1,000. The composite score is the arithmetic mean of the three pillar scores. The dictionary is positive-only by design: it captures the extent to which firms make sustainability claims, commitments, and disclosures, but does not penalise firms for discussing negative events or acknowledging shortcomings. Including negative keywords would conflate disclosure quality with disclosure quantity, which is not the object of our measurement.

Table A.1. Keyword dictionary by pillar and weight

The complete keyword list is reproduced below. Weight 2 entries indicate strong, explicit commitments or hard targets; weight 1 entries reflect softer disclosure, strategy, or policy language. † marks keywords that appear in more than one pillar.

Pillar	Weight 2 (strong commitments / hard targets)	Weight 1 (policy / strategy / disclosure language)
Environmental	net zero; net-zero; carbon neutral; carbon neutrality; science based target; science-based target; science based targets; science-based targets; sbti; paris aligned; paris agreement; emissions reduction target; decarbonisation target; decarbonization target; climate transition plan; transition plan; tcfd; task force on climate; esrs el	climate strategy; climate policy; climate related; climate-related; climate risk; climate risks; climate opportunity; climate opportunities; emissions target; emission target; emissions reduction; emission reduction; scope 1; scope 2; scope 3; carbon footprint; greenhouse gas; ghg emissions; renewable energy; energy efficiency; waste reduction; circular economy; biodiversity; water stewardship; environmental policy; environmental strategy; environmental target; sustainability strategy; sustainability target; double materiality†; csrd†; esrs†
Social	human rights policy; human rights due diligence; modern slavery statement; living wage; pay equity; gender pay gap target; zero fatalities; zero harm	human rights; diversity; inclusion; dei; employee wellbeing; health and safety; occupational health; community engagement; community investment; training and development; learning and development; employee engagement; social impact; equal opportunity; workplace safety; social target; social policy; stakeholder engagement; supplier code of conduct; responsible sourcing; ethical sourcing; accessibility; employee retention
Governance	anti-corruption; anti corruption; anti-bribery; anti bribery; board independence; whistleblower policy; code of conduct; tax transparency; double materiality†	corporate governance; governance; risk management; risk oversight; compliance; ethics; board diversity; independent director; internal controls; internal audit; external audit; transparency; disclosure; sustainability reporting; esg reporting; integrated reporting; gdpr; data privacy; cybersecurity; executive remuneration; shareholder rights; esrs†; csrd†

B. Variable Construction

B.1 Talk score formula and intuition

Formally, for firm i in year t and pillar $p \in \{E, S, G\}$, the talk score is

$$\text{talk}_{p,i,t} = \frac{\sum_{k \in D_p} w_k \cdot \mathbf{1}\{\text{match}_{k,i,t} \text{ and not negated}_{k,i,t}\}}{\text{wordcount}_{i,t}} \times 1,000, \quad (6)$$

where D_p is the set of keywords in pillar p , $w_k \in \{1, 2\}$ is the keyword weight, and the indicator function captures non-negated matches. The composite talk score is the arithmetic mean of the three pillar scores: $\text{talk}_{\text{composite},i,t} = \frac{1}{3}(\text{talk}_{E,i,t} + \text{talk}_{S,i,t} + \text{talk}_{G,i,t})$. Normalising by document length controls for the fact that longer reports would mechanically receive higher raw scores; without this step a firm publishing a 200-page integrated report would be measured as more sustainability-focused than one with a 60-page report regardless of substantive intensity.

A small number of keywords appear in more than one pillar. “Double materiality”,

“CSRD”, and “ESRS” are included in both the Environmental and Governance dictionaries because they inherently span both dimensions: double materiality is both an environmental reporting principle and a governance framework requirement under ESRS. Similarly, “sustainability reporting” and “integrated reporting” appear in the Governance dictionary despite having environmental relevance, reflecting their primary association with governance and disclosure practices. A single mention of “double materiality” contributes to both the E and G pillar scores but does not affect the composite score disproportionately, since the composite is the arithmetic mean of the three pillars.

B.2 Structured environmental talk proxy

Processing annual report text for 981 universe firms is not feasible. For the universe analysis we therefore construct a structured environmental talk proxy using three continuous LSEG scores that capture the same conceptual dimensions as the annual-report dictionary: whether the firm formally articulates environmental policies, sets quantitative targets, and assesses climate-related risk. The proxy is

$$\text{talk}_{\text{proxy E},i,t} = \frac{1}{3}(\text{PolicyEmissions}_{i,t} + \text{TargetsEmissions}_{i,t} + \text{ClimateRiskOpp}_{i,t}). \quad (7)$$

The three components parallel the sustainability-topic structure of the annual-report dictionary: Policy Emissions corresponds to policy-level vocabulary (climate policy, environmental policy), Targets Emissions corresponds to target-setting vocabulary (emissions reduction target, net zero), and Climate Change Risks/Opportunities corresponds to strategic assessment vocabulary (climate risk, transition plan). All three range from 0 to approximately 100 with meaningful cross-sectional variation (means 57.6, 57.2, 52.1; standard deviations 18.6, 32.8, 29.2). The proxy is environmental-only because the LSEG structured fields available for the social and governance dimensions do not offer a comparable set of commitment-focused scores.

We considered and excluded two binary alternatives. Waste Reduction Initiatives is coded “yes” for over 92% of firm-year observations, providing virtually no cross-sectional variation. Climate Change Strategy suffers from severe missingness in the early years of the sample (265 non-null observations in 2021 and 731 in 2022, out of approximately 1,000 firms), which would introduce a compositional break in the middle of the pre-period.

B.3 LSEG walk score construction

LSEG’s Environmental Pillar Score aggregates three category scores: emissions (commitment and effectiveness in reducing environmental emissions in production and operations), resource use (reductions in materials, energy, and water consumption), and innovation (development of environmental technologies and eco-designed products). Category weights vary by industry through LSEG’s materiality matrix; emissions receive a magnitude weight of 10 for metals and mining but 1 for banking, for example. Categories are percentile-ranked within each firm’s industry group, so the score captures relative performance among peers rather than absolute impact.

We use the base pillar scores rather than the controversy-adjusted ESGC scores. The ESGC discount uses severity weights that scale with market capitalisation: large firms receive a weight of 0.33, mid-cap firms 0.67, and small firms 1.0. Because our identification

strategy exploits a firm-size threshold, incorporating a size-dependent media penalty into the walk measure would introduce a mechanical confound.

B.4 Bid-ask spread and missing-value handling

The proportional bid-ask spread is constructed from daily LSEG bid and ask prices in EUR as $\text{Spread\%} = (\text{Ask} - \text{Bid}) / [(\text{Ask} + \text{Bid}) / 2] \times 100$. The bid-ask spread is a standard proxy for market liquidity and information asymmetry (Amihud, 2002). Including it as a control addresses the concern that differences in liquidity between treated and control firms could confound the estimated treatment effect, an issue raised in prior work on the relationship between ESG disclosure and market microstructure.

Some firms have missing ESG pillar scores for individual years. We do not fill missing values with zero, since a missing LSEG score indicates that LSEG did not assess the firm in that year, not that the firm scored zero. Instead we retain missing values as NaN and apply within-firm linear interpolation for gaps of at most one year. For a firm observed in years $t - 1$ and $t + 1$ but missing in year t , the interpolated value is the simple average of the adjacent observations. For missing values at the boundary of a firm’s observed period, we forward- or backward-fill within the firm. An imputation flag allows us to verify that results are not driven by imputed observations. In the final estimation sample, no surviving observation relies on interpolated walk scores; all interpolated values were dropped by the subsequent bandwidth and pre/post-period filters, so the main results are based entirely on directly observed LSEG pillar scores.

C. AI Disclosure

We used large language models (ChatGPT and Claude) during this project for the following tasks: proof-reading and copy-editing of draft text; debugging Python code for annual-report text extraction, keyword matching, and negation detection; and spot-checking the consistency of coefficient reporting between code output and the manuscript. All substantive research choices, including the research question, identification strategy, sample construction, dictionary design, and interpretation of results, were made by the authors. No model output was incorporated into the manuscript without human review and editing.