Discipline or display?

CSR engagement and earnings management in EU-listed firms before and after the NFRD

Ilona Chaiko
Baptiste Richaudeau

Master Thesis

Stockholm School of Economics

2026

Discipline or display? CSR engagement and earnings management in EU-listed firms before and after the NFRD

Abstract

This study examines whether Corporate Social Responsibility (CSR) engagement constrains earnings management among European listed firms over the 2010-2025 period. Using a panel of 1,944 firms across the 27 EU member states, we test two related questions: whether higher Environmental, Social, and Governance (ESG) performance is associated with lower magnitudes of accrual-based and real-activities earnings management, and whether the 2017 Non-Financial Reporting Directive (NFRD), exploited as a quasi-natural experiment within a two-stage least-squares design, exerts a disciplining causal effect on financial reporting quality. The evidence is predominantly null. The baseline cross-sectional association vanishes once firm-level controls and saturated fixed effects are introduced. A marginally negative accrual-based effect emerges only under a joint-substitution specification and concentrates in the Environmental pillar. Although the NFRD produced a large first-stage increase in ESG performance among ex ante low-CSR firms, the directive did not translate into a detectable reduction in either manipulation channel, and the predicted accrual-to-real substitution does not materialise. We interpret these null results as reflecting (i) the civil-law institutional context in which reporting discipline may already be strong, (ii) the dilution of CSR signalling under universal mandatory disclosure, and (iii) the limitations of linear cross-sectional ESG proxies when the CSR-earnings management relationship is non-monotonic.

Keywords:
CSR performance, Earnings Management, Accrual-Based Earnings Management (AEM), Real Earnings Management (REM), NFRD

Authors:
Ilona Chaiko (42773)
Baptiste Richaudeau (42763)

Tutor:
Irina Gazizova, Assistant Professor, Department of Accounting

Master Thesis
Master Program in Accounting, Valuation and Financial Management
Stockholm School of Economics

Acknowledgements

We are deeply grateful to our supervisor, Assistant Professor Irina Gazizova from the Department of Accounting, for her continuous support throughout the thesis work. Her guidance, insightful feedback, and valuable comments, in particular regarding the empirical design and statistical methodology, have been instrumental to the completion of this study.

Stockholm
Ilona Chaiko & Baptiste Richaudeau

Table of contents

1. Introduction

Whether a firm's commitment to social and environmental conduct extends to the integrity of its financial reporting is one of the central empirical questions of contemporary accounting research. The growing stakeholder demand for transparency along non-financial dimensions, the proliferation of Environmental, Social, and Governance (ESG) information and ratings, and the regulatory turn toward mandatory sustainability disclosure have together placed this question at the intersection of accounting, governance, and public policy. The question itself, however, remains contested. The international empirical evidence is fragmented across institutional settings, the theoretical mechanisms point in opposite directions, and the cross-sectional designs that dominate the existing literature struggle to support the causal interpretations that policy-makers and stakeholders increasingly demand. This study revisits the question within the European Union, exploiting a regulatory event that offers identification possibilities largely absent from prior work, and integrating two stages of the CSR-reporting chain that the literature has so far examined in isolation.

Three features of the present empirical setting motivate a renewed examination of the relationship. The first is institutional. The European Union constitutes a fairly different setting from the United States economy in which much of the foundational evidence on Corporate Social Responsibility (CSR) and financial reporting was developed, and the conditions under which any cross-domain ethical mechanism might operate differ accordingly. The second is regulatory. The 2017 implementation of the Non-Financial Reporting Directive (Directive 2014/95/EU; hereafter referred to as the NFRD) transformed CSR disclosure for large European public-interest entities from a voluntary practice into a binding legal obligation, providing a quasi-exogenous shock whose compliance bite varied predictably across firms with different pre-existing levels of CSR engagement. To our knowledge, this study is the first to use the NFRD as an identifying shock for the CSR-earnings management relationship; prior European work has either pooled pre- and post-mandate observations or relied on voluntary-disclosure proxies, leaving the causal claims of the moral-track literature largely untested in a mandatory-regime setting. The third is methodological. The two principal channels through which managers manipulate reported earnings, namely accrual-based discretion and real-activities manipulation, may respond differently to regulatory pressure on CSR disclosure, and aggregate ESG ratings may obscure heterogeneity across the Environmental, Social, and Governance components.

This study examines the CSR-earnings management relationship within a panel of 1,944 European listed firms spanning the 2010 to 2025 period and the twenty-seven member states of the European Union. The empirical analysis is organised around two interconnected research questions. The first considers whether aggregate ESG performance is associated with lower magnitudes of accrual-based and real-activities earnings management in the European cross-section, thereby adjudicating between two competing theoretical predictions: that CSR commitment reflects an ethical orientation that constrains opportunistic reporting, and that CSR engagement provides a reputational cover under which opportunistic reporting may proceed unchecked. The second question exploits the 2017 NFRD shock within a two-stage least-

squares framework to assess whether mandatory CSR engagement causally constrains earnings management, or whether the heightened regulatory scrutiny instead displaces opportunism from accrual-based manipulation toward less detectable real-activities manipulation. In doing so, the study estimates the full chain from mandatory disclosure, through CSR performance, to financial reporting behaviour as an integrated mechanism, rather than examining its component links separately as the existing literature has tended to do.

The contribution of the study is therefore fourfold. First, it provides one of the first causal estimates of how a CSR-disclosure mandate affects financial reporting quality, moving the European literature beyond the correlational and pooled designs that have so far dominated. Second, it tests the transmission from disclosure regulation to reporting outcomes as a single integrated chain, an exercise the prior literature has discussed conceptually but rarely executed empirically. Third, it extends the empirical evidence to a civil-law institutional environment whose architecture renders direct extrapolation from the United States problematic, and uses the resulting (largely null) findings to delineate the boundary conditions of the moral-track hypothesis. Fourth, it disaggregates both the dependent and the independent variables in order to identify the specific channels through which any CSR effect operates, allowing the analysis to detect within-construct heterogeneity that aggregate-level studies cannot.

The empirical evidence is predominantly null. The bivariate correlations between aggregate ESG and the two earnings management proxies, while negative and statistically significant in the raw data, do not survive the introduction of firm-level controls and saturated fixed effects. A marginally negative association on accrual-based manipulation emerges only under a joint-substitution specification, with the Environmental pillar accounting for the entirety of this effect. The NFRD generates a substantial first-stage increase in ESG performance among treated firms, yet this exogenous variation does not translate into a detectable reduction in either manipulation channel, and the substitution dynamic anticipated under heightened regulatory scrutiny is not observed. In the European cross-section, the two channels behave as complements rather than substitutes. The conditions under which these results have been obtained, and the institutional, regulatory, structural, and measurement factors that plausibly explain them, are developed at length in the Sections that follow.

2. Literature Review

This Section reviews the theoretical and empirical literature surrounding the relationship between CSR and financial reporting.

2.1 Introduction to the CSR-Earnings Management Relationship

As stakeholders increasingly expect openness beyond financial measures, accounting research has focused on the relationship between Corporate Social Responsibility (CSR) and the quality of financial reporting. Carroll (1979) asserts that the framework, which is broken down into four categories (economic, legal, ethical, and discretionary), can be used to describe a company's social duty to the society in which it operates. Whether a company's dedication to

these obligations translates into a commitment to reporting the truth in financial statements is at the center of this debate.

This study asks if organizations with strong CSR performance also act morally to limit earnings management (EM), building on the fundamental question of Kim et al. (2012) based on American (US) businesses. Global evidence is still dispersed despite the ethical connection put forth by Kim et al. (2012). Through their global study conducted as early as 2008, Chih et al. brought attention to the erratic relationship between CSR and EM, which is heavily influenced by the institutional framework. Liang and Renneboog (2017) show that a firm's CSR engagement is essentially rooted in its nation's legal origin and that civil law jurisdictions, like those in the European Union, foster a higher baseline for social responsibility as a matter of institutional compliance, further solidify this institutional perspective. Extending this complexity, Elrazaz and Aljifri (2026) use nonlinear approaches to show that the ESG-EM relationship is frequently non-monotonic, implying that ESG's capacity to limit opportunism might only be useful within specific performance thresholds.

Determining the direction of causality, however, continues to be a problem in this research. Reverse causality, in which companies with cleaner financial books invest more heavily in ESG to signal discipline, or unobserved firm characteristics, like managerial ethics or stakeholder orientation, which simultaneously elevate ESG scores and constrain earnings management, could also be the cause of the observed association between high CSR performance and higher-quality financial reporting. The quasi-experimental portion of this study is motivated by the need for an identification technique that isolates an exogenous source of variation in CSR performance in order to disentangle various channels.

This review navigates the theoretical tensions that explain these mixed results, structured around two interconnected research questions. The first examines the baseline association between CSR performance and earnings management, testing whether ethical commitment constrains financial manipulation (H1.a) or whether CSR serves as a reputational shield for opportunistic behavior (H1.b). The second exploits the EU Non-Financial Reporting Directive (NFRD) as a quasi-exogenous regulatory shock to move closer to causal identification. This transition highlights a critical shift from voluntary disclosure, often viewed as an ethical signal, to mandatory regimes where reporting is driven by regulatory discipline (Xue et al., 2025). Under this mandatory framework, we test whether the directive is associated with a genuine decrease in the overall magnitude of earnings management through the internalization of ethical conduct (H2.a), or whether it merely incentivized a substitution effect, whereby heightened scrutiny from analysts and auditors makes Accrual-Based Earnings Management (AEM) increasingly risky and detectable (He et al., 2022), forcing managers to pivot toward Real Activities Manipulation (REM). Because REM is embedded in operational decisions like cutting R&D or overproducing inventory, it is significantly harder for outsiders to distinguish from legitimate business strategy (Zang, 2012; Bozzolan et al., 2015). Consequently, the NFRD analysis tests whether mandatory reporting fostered genuine transparency or merely forced opportunistic behavior into less visible operational channels (H2.b) (Zang, 2012; Xue et al., 2025).

2.2 Theoretical Perspectives

This section establishes the conceptual boundaries of CSR and EM to contextualize how social performance interacts with financial reporting behaviors.

2.2.1 Theoretical Perspectives on CSR

Garriga and Melé (2004) divide the existing studies into four categories to explain the reasons behind CSR: instrumental, integrative, political, and ethical. The literature is mainly unsatisfactory since these frameworks have produced conflicting empirical findings about the connection between CSR and earnings management. According to some research, managers may use CSR as an opportunistic smoke screen to cover up wrongdoing (Petrovits, 2006; Prior et al., 2008). On the other hand, some find a negative correlation, the moral track, where managers are motivated to generate better financial reports due to ethical duties and reputational concerns (Jones, 1995; Fombrun and Shanley, 1990; Kim et al., 2012).

2.2.2 Theoretical Perspectives on Earnings Management

Assessing financial quality requires looking beyond accounting estimates. A critical distinction in this thesis is the form that earnings management takes. Managers generally choose between two methods:

Accrual-Based Earnings Management (AEM) involves the opportunistic application of accounting discretion and subjective judgment within the accrual principle to manipulate reported net income and modify estimates and choices in financial reports. Unlike changes in real business operations, AEM focuses on manipulating the soft components of financial statements (such as loss provisions, depreciation estimates, and revenue timing) to shift earnings across periods without affecting current-period cash flows (Healy & Wahlen, 1999). Dechow and Skinner (2000) and Kothari et al. (2005) argue that distinguishing legitimate cash-flow smoothing from abusive AEM is challenging because managerial intent cannot be directly observed. This difficulty has a significant methodological implication: high-performing firms inherently generate larger accruals due to their economic growth, which means that models failing to control for firm performance may erroneously attribute legitimate growth-driven accruals to opportunistic manipulation. This issue provides the theoretical justification for using performance-adjusted accrual models, like the one proposed by Kothari et al. (2005), which include return on assets as a matching variable to ensure that discretionary accruals are measured in comparison to firms with similar profitability profiles.

Prior research describes AEM as a manager's first choice despite this detection challenge because it avoids the immediate costs to company value associated with operational distortions (Zang, 2012). Kasznik (1999) shows that managers employ income-increasing discretionary accruals to guide reported earnings toward their own voluntary projections in order to reduce potential legal and reputational consequences. This is another example of this strategic use of accounting discretion. This study concentrates on the total amount of discretionary accruals, even if managerial intent can influence earnings in both positive and negative ways. This

method, which is in line with Kim et al. (2012), views any notable departure from normal accruals, regardless of direction, as a decrease in the quality of financial reporting since it distorts the firm's actual underlying economic performance.

However, its application is strictly limited by the reversal property of accruals (the accounting reality that an inflation of earnings today must be offset in the future), which creates a cumulative constraint on the balance sheet known as a lack of accounting slack (Barton & Simko, 2002). As Dechow et al. (2012) demonstrate, traditional models struggle to separate genuine growth from manipulation. Consequently, explicitly modeling this self-correcting reversal of accruals is crucial for reliably detecting AEM. Because of these mounting balance sheet constraints, coupled with increased audit quality and diminished GAAP flexibility, managers are eventually forced to abandon AEM in favor of alternative manipulation routes (Zang, 2012).

When managers depart from typical business procedures in order to deceive stakeholders into thinking that specific financial reporting objectives have been met through routine operations, this is known as Real Activities Manipulation (REM) (Roychowdhury, 2006). In contrast to accrual-based adjustments, REM entails deliberate modifications to the timing or structure of physical activities. Examples include overproducing to reduce unit costs by spreading fixed overhead, accelerating sales through excessive price discounts, and cutting discretionary expenditures in R&D or advertising. Because they compromise long-term investment and future cash flows, these operational distortions are intrinsically more expensive than AEM. However, because they do not violate GAAP, they are much more difficult for external parties to detect (Graham et al., 2005; Roychowdhury, 2006).

Notably, managers often treat AEM and REM as strategic substitutes, opting for operational distortions when the costs or risks of accounting manipulation become too high (Cohen and Zarowin, 2010; Badertscher, 2011; Zang, 2012).

As documented by Cohen et al. (2008), REM is inherently more difficult for external parties to identify because it does not violate GAAP and is often camouflaged as normal business decisions. Because of this diminished detection risk, managers typically prefer REM over AEM, especially when firms face heightened regulatory scrutiny or have exhausted their balance sheet's capacity for further accrual manipulation.

2.3 Syntheses and Hypotheses

This section synthesizes the preceding theories and empirical evidence to develop the study's hypotheses.

2.3.1 The Transparent Financial Reporting Hypothesis as a foundation of the negative association (The Moral Track)

The moral track theory, which has its roots in Stakeholder Theory (Freeman, 1998), asserts that CSR is motivated by a moral and normative duty to stakeholders. According to this theory, managers who place a high priority on social responsibility naturally cherish honesty and moral

behavior, fostering a broad ethical culture that serves as an internal governance framework. Ricapito (2024) affirms that ESG performance in the Italian setting considerably restricts earnings management, with the social and governance pillars playing a major role in supporting reporting quality, based on recent data from the European landscape. According to Kim et al. (2012), a company's dedication to moral behavior is inseparable, implying that the moral integrity that propels social performance inevitably permeates its financial reporting. This moral track is supported by ethical, political, and integrative views, which argue that socially responsible managers are inherently driven by a commitment to honesty and transparency. Consequently, integrity is viewed as a holistic attribute where high-quality financial reporting and substantive CSR performance mutually reinforce one another. Managers in such firms view the manipulation of earnings as a violation of their fundamental duty to stakeholders (Gras-Gil et al., 2016). This is further evidenced by Gavana et al. (2022), who find that high ESG performance serves as a critical moderating variable that inhibits the use of related party transactions for opportunistic earnings management, thereby reinforcing internal governance frameworks.

Moreover, CSR plays a crucial role in building and maintaining a company's reputation (Fombrun & Shanley, 1990; Verschoor, 1998). Firms with high CSR performance have substantial reputational capital at stake, which deters managers from engaging in socially unacceptable accounting practices like earnings management that could erode stakeholder trust (Godfrey, 2005). The significance of these reputational concerns is further emphasized by Kasznik (1999), who shows that the desire to avoid damaging the company's reputation for accuracy is a key driver for managers to use discretionary accruals to meet publicly disclosed financial targets.

External supervision channels supplement this innate ethical urge. By giving outsiders operational data to evaluate the reliability of earnings, high-quality ESG engagement serves as a positive signal that lessens information asymmetry (Dhaliwal et al., 2011; Xue et al., 2025). Financial analysts' increased scrutiny of superior social performance serves as a supervisory tool that effectively limits a manager's space for opportunistic manipulation (He et al., 2022; Sun et al., 2024). This relationship is particularly potent when external assurance is present, as high-quality ESG engagement provides outsiders with the operational data necessary to assess the credibility of reported earnings (Dhaliwal et al., 2011; Xue et al., 2025). Ultimately, this alignment of ethical obligation and external monitoring effectively constrains opportunistic behavior, leading to a predicted negative association between CSR and EM. The convergence of these internal and external channels allows us to hypothesize the following:

H1.a: Higher ESG scores are associated with a lower magnitude of earnings management (higher financial reporting quality) across both AEM and REM channels.

2.3.2 The Opportunistic Financial Reporting Hypothesis as a foundation of the positive association (Moral Licensing)

Agency Theory (Jensen & Meckling, 1976) and instrumental viewpoints, in contrast to the moral perspective, contend that CSR can be used for managerial self-interest or as a calculated

tactic to build reputational capital. According to this viewpoint, corporate social responsibility (CSR) is a managerial requirement used to achieve individual career ambitions, forward personal agendas, or avoid criticism rather than a show of virtue. According to this viewpoint, self-serving managers use corporate social responsibility (CSR) to build a favorable public image, which subsequently acts as a smoke screen to conceal opportunistic actions and corporate wrongdoing (Prior et al., 2008; Xue et al., 2025). Certain philanthropic methods, such corporate-sponsored foundations, have occasionally been utilized as off-balance sheet reserves to make discretionary funding decisions that increase income, especially when businesses have strong equity market incentives to fulfill profit targets (Petrovits, 2006). As a result, CSR reporting can serve as a tool for managerial discretion, enabling executives to present a responsible image while deceiving stakeholders about the firm's actual worth and financial success. As Dechow and Skinner (2000) point out, managers have strong incentives to meet basic earnings benchmarks in order to protect stock valuations, which encourages them to manipulate financial reports, especially when corporate governance structures are weak. This opportunistic behavior is frequently reinforced by intense capital market pressures.

The Moral Licensing Hypothesis provides the theoretical foundation for this opportunistic behavior. Ormiston and Wong (2013) demonstrate that managers often derive a psychological license from their good deeds in the CSR domain, which they subsequently use as moral credit to justify or conceal opportunistic behavior in other areas, such as the financial domain. In this context, CSR activities serve as a mechanism of self-impression management designed to attract stakeholders without substantially implementing sustainable transformation goals. Wadhaad et al. (2023) find empirical evidence supporting this effect, especially in financially distressed firms, where managers use high ESG scores to disguise declining performance and maintain access to capital. Likewise, Moharram et al. (2026) show that the potential performance benefits of ESG disclosure are diminished when managers simultaneously engage in AEM or REM, indicating that the market eventually penalizes firms that attempt to use disclosure as a cover for operational manipulation. Managers may internalize the negative consequences of their behavior and attempt to conceal them by adopting extensive ESG disclosures to create an appearance of transparency.

This form of reputation insurance effectively grants managers a license to operate opportunistically. Under the Moral Licensing hypothesis, this manifests as a higher intensity of discretionary behavior, where the absolute distance between reported and normal earnings increases as managers exploit their perceived ethical capital. In these settings, mandatory disclosures may even exacerbate the issue if firms view reporting as a mere compliance task rather than a substantive commitment (Xue et al., 2025). Under these conditions, the pursuit of organizational economic egoism drives a positive association between social performance and financial manipulation. Given these motivations, the alternative relationship is hypothesized as follows:

H1.b: Higher ESG scores are associated with a higher magnitude of earnings management (lower financial reporting quality) across both AEM and REM channels.

2.3.3 The NFRD as a Regulatory Shock: Ethical Internalization and External Monitoring

The literature finds external processes that link required disclosure to the quality of financial reporting, going beyond the basic correlation between CSR and earnings management. According to Kim et al. (2012) and the moral track hypothesis, there is a strong association between ethical managers' voluntary reporting of CSR data and their later reporting of more transparent and high-quality financial data. CSR disclosure serves as a reliable signaling tool in a voluntary situation, enabling genuinely moral managers to set themselves apart from opportunistic colleagues (Dhaliwal et al., 2011; He et al., 2022). But this dynamic is drastically changed by the transition from optional to required non-financial reporting (like the NFRD in Europe). Because reporting is now a universal legal requirement rather than a discretionary choice, this signaling effect is neutralized; stakeholders can no longer rely on the mere existence of a CSR report to identify an ethical manager (Xue et al., 2025). As a result, the regulatory domain becomes the driving factor behind transparency instead of the discretionary domain (Christensen et al., 2021).

However, mandated reporting has an effect that goes beyond simple external discipline. The literature indicates a deeper, internalization pathway by which obligatory CSR participation may change executive behavior, even while increased scrutiny from analysts, auditors, and institutional investors clearly raises the expected penalties of financial malfeasance. In their thorough analysis, Christensen et al. (2021) confirm that making CSR reporting mandatory can significantly alter the company's internal information environment and decision-making procedures in addition to giving outsiders improved monitoring tools. The mandate may incorporate ethical issues into the company's regular operations and strategic planning by requiring managers to rigorously gather, assess, and report on environmental, social, and governance factors. Fiechter et al. (2022) provide direct evidence that the NFRD mandate did not merely alter what firms report, but actually forced them to change their real operational behaviors, suggesting that the mandatory engagement with sustainability topics can catalyze genuine shifts in corporate conduct rather than purely cosmetic compliance. In this sense, the institutionalization of reporting obligations may lead managers to internalize more ethical conduct in their financial reporting as a byproduct of their deepened engagement with stakeholder-oriented governance.

To comprehend the external dimension of this mechanism, we draw a parallel to Cohen et al. (2008), who established a clear connection between the implementation of mandatory reporting regimes and a subsequent improvement in the quality of financial information. Following the enactment of the Sarbanes-Oxley Act (SOX), the U.S. market experienced a structural decrease in accounting manipulation due to unprecedented regulatory oversight. Similarly, the NFRD introduces a heightened level of scrutiny by specifically targeting public-interest European entities with an average of more than 500 employees. For these major economic players, the directive explicitly requires reporting on environmental, social and employee matters, respect for human rights, anti-corruption and bribery matters. According to He et al. (2022), this mandatory influx of non-financial data triggers increased attention from financial analysts, brokerage firms, and institutional investors. This influx of professional observers creates a powerful scrutiny effect, equipping the market with a more comprehensive toolkit to assess

firm fundamentals (Dhaliwal et al., 2011). In particular, detailed ESG disclosures regarding workforce, supply chains, or governance serve as an informational bridge for outsiders, helping them determine whether a firm's reported financial growth is structurally plausible (Dhaliwal et al., 2011). By cross-referencing these operational realities against accounting figures, external monitors can more readily identify inconsistencies. Indeed, Xue et al. (2025) demonstrate that this reduction in information asymmetry is the primary channel through which ESG transparency influences earnings quality, as it significantly limits the black box of managerial discretion that typically enables opportunistic reporting.

Crucially, the literature demonstrates that the NFRD's substantive bite is not uniform across all firms but concentrates heterogeneously among those with the weakest pre-existing CSR engagement. Fiechter et al. (2022) establish that the real effects of the NFRD on CSR activities are concentrated in firms with low pre-directive levels of both CSR reporting and CSR performance, firms with the greatest scope for behavioral change under mandatory disclosure. Grewal et al. (2019) come to a consistent conclusion: firms with bad pre-directive ESG performance and transparency are the focus of equity market reactions to the NFRD's enactment, whereas firms with high pre-directive performance see favorable reactions. This heterogeneity is crucial to the study's empirical design because firms with the weakest pre-mandate CSR positioning are disproportionately subject to regulatory pressure, making them the population for which the mandate is most likely to result in a legally binding change in both CSR behavior and, consequently, financial reporting conduct. The cross-sectional variation required to separate the influence of the NFRD from other confounding factors affecting the European market over the same period is provided by the identification of this differential exposure based on pre-treatment CSR levels.

The convergence of these internal and external channels, the internalization of ethical conduct through mandatory engagement with sustainability reporting, combined with the intensified external monitoring that the resulting disclosures enable, suggests that the NFRD is associated with a disciplining effect on financial reporting quality. Following this logic, we hypothesize:

H2.a: The transition from voluntary to mandatory ESG disclosure under the NFRD generates a significant decrease in the overall magnitude of earnings management, as the institutionalization of ethical, environmental, and social reporting obligations leads managers to internalize more ethical conduct in their financial reporting.

2.3.4 The Strategic Substitution Hypothesis as a consequence of mandatory scrutiny

While the combined internalization and scrutiny effects induced by mandatory reporting may successfully reduce highly visible accounting manipulation, contemporary research suggests that this same regulatory pressure simultaneously drives a strategic substitution effect. The critical insight, established in the post-SOX literature by Cohen et al. (2008), is that as the cost of accounting manipulation increases due to regulatory oversight, managers do not necessarily cease manipulation; rather, they pivot toward less detectable methods.

Because AEM leaves a clear audit trail and is severely constrained by the cumulative absence of accounting slack on the balance sheet, managers try to avoid it as obligatory CSR reporting and external monitoring create reputational pressure (Barton & Simko, 2002; Zang, 2012). Managers turn to Real Activities Manipulation (REM) when the cost of accrual manipulation increases as a result of such regulatory supervision (Cohen et al., 2008). Opportunistic conduct can continue in a less obvious form since REM is strongly ingrained in operational decisions, making it much more difficult for outsiders to discern from legitimate company strategy (Roychowdhury, 2006; Bozzolan et al., 2015). This managerial preference is not merely theoretical; Graham et al. (2005) documented through extensive executive surveys that financial officers overwhelmingly prefer to utilize real economic actions rather than accounting choices to meet financial targets, specifically to evade the severe consequences of auditor scrutiny. According to Zang (2012) and Cohen and Zarowin (2010), this trade-off is a direct function of the firm's specific environment: when the cost of accrual manipulation rises due to heightened scrutiny, managers strategically shift toward executing real manipulations during the fiscal year.

This substitution dynamic following a mandatory CSR shock is empirically supported by Fiechter et al. (2022), who provide evidence that the NFRD mandate in the European Union did not merely alter what firms report, but actually forced them to change their real operational behaviors due to the sudden transparency imposed on their standard accounting choices. While Chen et al. (2018) document that such mandates can force a trade-off between social performance and financial profitability, the pressure to maintain these financial figures under heightened scrutiny often drives managers toward less detectable manipulation. According to Zang (2012) and Cohen and Zarowin (2010), when the cost of accrual manipulation rises due to mandatory oversight, managers strategically shift toward executing real manipulations (REM), as these are inherently harder for auditors to distinguish from legitimate business strategy (Roychowdhury, 2006).

Consequently, under the NFRD, while the financial books (AEM) become subject to greater scrutiny by auditors and stakeholders, operational decisions (REM) remain inherently more challenging to monitor. Drawing on the evidence from Cohen et al. (2008) regarding the post-SOX substitution effect, alongside the contemporary findings of Fiechter et al. (2022), this suggests that increased regulatory pressure may not completely eliminate opportunism, but instead force it into the side window of operational distortions, which are often more detrimental to long-term firm value (Bozzolan et al., 2015). The external monitoring pressure generated by the NFRD, through analyst scrutiny, auditor attention, and the informational bridge provided by mandatory disclosures, disproportionately increases the detection risk and reputational cost of accrual-based manipulation, while leaving the operational domain relatively unmonitored. This asymmetry in monitoring intensity creates a rational incentive for managers to shift toward less detectable forms of opportunistic reporting. We hypothesize that this regulatory pressure prompts a strategic shift:

H2.b: The implementation of mandatory ESG disclosure under the NFRD generates a substitution effect in earnings management behavior, whereby firms reduce their reliance on accrual-based earnings management (AEM) and increase their use of real activities manipulation (REM), as external monitoring pressure incentivizes a shift toward less detectable forms of opportunistic reporting.

3. Methodology

This Section outlines the research design and statistical models used to examine the relationship between Corporate Social Responsibility (CSR) and earnings management. Section 3.1 describes the construction of the two earnings management proxies, the CSR measure, the control vector, and the outlier-treatment procedure. Section 3.2 presents the OLS framework that evaluates the conditional CSR-earnings management association. Section 3.3 presents the instrumental-variables design that exploits the 2017 EU Non-Financial Reporting Directive (NFRD) as a quasi-natural experiment to move closer to causal interpretation.

3.1 Measurement of Earnings Management

Following the prior literature, we use two complementary proxies: accrual-based earnings management (AEM) and real earnings management (REM).

3.1.1 Accrual-Based Earnings Management (AEM)

AEM captures the opportunistic use of accounting discretion to alter reported net income without affecting current-period cash flows (Healy & Wahlen, 1999). Because high-performing firms generate larger accruals as a function of economic growth (Dechow et al., 2012), models that fail to control for performance risk confuse growth-driven accruals with manipulation (Dechow & Skinner, 2000). We therefore estimate discretionary accruals using the performance-adjusted Modified Jones model of Kothari et al. (2005), cross-sectionally for each NAICS subsector-year cell with at least ten observations. Total accruals are defined as TA = Net Income - CFO and scaled by lagged total assets A_{t-1}.

Step 1 (parameter estimation)**.** We estimate the cross-sectional regression:

$$TA_{i,t}/A_{i,t-1} = \beta_0 + \beta_1(1/A_{i,t-1}) + \beta_2(\Delta REV_{i,t}/A_{i,t-1}) + \beta_3(PPE_{i,t}/A_{i,t-1}\ ) + \beta_4 ROA_{i,t-1} + \varepsilon_{i,t} \quad (1)$$

where $\Delta REV_{i,t}$ is the change in sales, PPE is gross property, plant and equipment, and $ROA_{i,t-1}$ is lagged return on assets. Lagged rather than contemporaneous ROA is used (per Kothari et al., 2005) to avoid mechanical correlation between the dependent variable and a regressor.

Step 2 (residual calculation). Following Kim et al. (2012), we apply the estimated industry coefficients to compute non-discretionary accruals while substituting ($\Delta REV_{i,t} - \Delta REC_{i,t}$) for $\Delta REV_{i,t}$ to flag revenue manipulation through credit sales:

$$NDA_{i,t} = \beta_0 + \beta_1(1/A_{i,t-1}) + \beta_2((\Delta REV_{i,t} - \Delta REC_{i,t})/A_{i,t-1}) + \beta_3(PPE_{i,t}/A_{i,t-1}) + \beta_4 ROA_{i,t-1} + \varepsilon_{i,t} \quad (2)$$

Discretionary accruals are then *DA_KM* = TA/A - NDA. Following Kim et al. (2012), our multivariate measure is the absolute value of these residuals, |*DA_KM*|, on the grounds that any significant deviation from normal accruals, regardless of direction, represents a distortion of underlying economic performance.

3.1.2 Real Earnings Management (REM)

Real earnings management captures operational distortions used to meet financial targets without violating GAAP. Following Roychowdhury (2006), we focus on three channels: sales acceleration via aggressive discounts or lenient credit (which depresses cash flow per unit), overproduction (which spreads fixed costs over a larger inventory base and lowers reported COGS), and cuts to discretionary spending on R&D and SG&A. These distortions are operationally costlier than AEM because they impair future cash flows, but are harder for outsiders to detect (Graham et al., 2005). As with AEM, all three normal-operations regressions are estimated cross-sectionally for each NAICS subsector-year cell with at least ten observations.

Normal cash flow from operations, normal production costs, and normal discretionary expenses are estimated from

$$CFO_{i,t}/A_{i,t-1} = \gamma_0 + \gamma_1(1/A_{i,t-1}) + \gamma_2(REV_{i,t}/A_{i,t-1}) + \gamma_3(\Delta REV_{i,t}/A_{i,t-1}) + \varepsilon_{i,t} \quad (3)$$

$$PROD_{i,t}/A_{i,t-1} = \delta_0 + \delta_1(1/A_{i,t-1}) + \delta_2(REV_{i,t}/A_{i,t-1}) + \delta_3(\Delta REV_{i,t}/A_{i,t-1}) + \delta_4(\Delta REV_{i,t-1}/A_{i,t-1}) + \varepsilon_{i,t} \quad (4)$$

$$DISEXP_{i,t}/A_{i,t-1} = \theta_0 + \theta_1(1/A_{i,t-1}) + \theta_2(REV_{i,t-1}/A_{i,t-1}) + \varepsilon_{i,t} \quad (5)$$

where PROD is defined as COGS plus the change in inventory and DISEXP as the sum of R&D and SG&A. Following Roychowdhury (2006), missing R&D values are coded as zero when SG&A is reported, to avoid biasing the sample toward innovation-intensive industries. Abnormal CFO, PROD, and DISEXP are the firm-specific residuals from equations (3)-(5). Income-increasing real manipulation produces unusually low CFO, unusually high PROD, and unusually low DISEXP. We therefore construct an aggregate proxy whose direction is consistent across all three channels:

$$Total\ REM_{i,t} = Abn\ PROD_{i,t} - Abn\ CFO_{i,t} - Abn\ DISEXP_{i,t} \quad (6)$$

A higher Total_REM score signals a greater likelihood of severe real-activity manipulation. The cross-sectional cell-size threshold of ten observations is verified through a sensitivity analysis (Section 3.3.7.5) that re-estimates the model with stricter thresholds of 15 and 20.

3.1.3 CSR Performance

We proxy CSR performance with the aggregate ESG score from LSEG Workspace, a 0-100 metric calculated from public sources across ten categories (resource use, emissions, workforce, human rights, management quality, and others) and weighted by industry-specific materiality (LSEG, 2024; Liang and Renneboog, 2017). In addition to the aggregate score, we separately analyse the three constituent pillars (Environmental, Social, Governance). The aggregate score offers a broad view of a firm's ethical standing but can mask specific drivers of managerial behaviour. Pillar-level analysis identifies which dimension of sustainability, employee relations (Social) or internal oversight (Governance), for example, most constrains different types of reporting manipulation.

3.1.4 Control Variables

We include controls shown by prior work to influence reporting incentives. **Size** is log market capitalisation. Larger firms face greater scrutiny and have weaker manipulation incentives (Watts & Zimmerman, 1990). **MB** (market-to-book) captures growth-related capital-market pressure to meet earnings benchmarks (Prior et al., 2008). **ROA** is included for performance matching following Kothari et al. (2005). **Lev** (long-term debt / total assets) captures debt-covenant pressure (Watts & Zimmerman, 1990). **Log_Age** = log(1 + firm age) controls for the more stable accrual processes of mature firms. The unit-shift transformation ensures the variable is well-defined for firms with zero or one year of history. **CFO_scaled** is operating cash flow scaled by lagged total assets, accounting for the mechanical accrual-cash-flow relationship. External monitoring is captured by **Analyst_Cov** (number of analysts following the firm) and **Big4** (binary indicator for a Big-4 auditor) (Francis et al., 1999). Following standard practice, we treat missing auditor identification as informative and set Big4 to NA rather than coding non-Big4. The implications are discussed in Section 3.3.7.4. Finally, following Zang (2012) and Cohen et al. (2008), we cross-control for the alternative manipulation proxy in our joint specifications: AEM models include Total_REM and REM models include |DA_KM| to ensure that the observed CSR effect is not merely a substitution artefact. All regressions include NAICS-subsector and year fixed effects.

3.1.5 Outlier Treatment

To limit the influence of extreme values on coefficient estimates, all continuous variables are winsorized at the 1st and 99th percentiles. Winsorization preserves rank ordering while bounding the impact of data-entry errors and genuine outliers (Cohen et al., 2008; Roychowdhury, 2006). The procedure is applied after the construction of all earnings management measures and control variables and before the multivariate analysis.

3.2 Empirical Models

Our empirical strategy proceeds in two complementary phases corresponding to the two research questions established in Section 2. The first phase, in this section, evaluates the conditional CSR-EM association via OLS in order to discriminate between the Transparent

Financial Reporting Hypothesis (H1.a, predicting a negative association) and the Opportunistic / Moral Licensing Hypothesis (H1.b, predicting a positive association). The second phase, in Section 3.3, evaluates the causal effect of mandatory CSR disclosure on EM by exploiting the NFRD as a quasi-exogenous regulatory shock, discriminating between the Mandatory Disclosure Discipline Hypothesis (H2.a) and the Strategic Substitution Hypothesis (H2.b). A complete summary of all regression equations is provided in Appendix A.

3.2.1 Stepwise Regression Framework

The OLS analysis proceeds in four steps. Step 1 is a large-sample sensitivity analysis on a reduced specification that excludes MB and Lev (the two controls with the highest data attrition) to validate that our core findings are not driven by sample-selection effects from those variables. Step 2 estimates the primary aggregate models on the strictly filtered sample with the full vector of controls. Step 3 disaggregates the analysis in two directions: it replaces the aggregate ESG score with individual Environmental, Social, and Governance pillar scores to identify which dimension of CSR most constrains manipulation (motivated by Ricapito (2024), who finds Social and Governance dominant in the European setting, and Gavana et al. (2022), who document ESG as a moderator of opportunistic behaviour), and separately decomposes the Total_REM effect into its three operational channels. Step 4 is the joint estimation discussed in 3.2.2. All the steps in detail are provided in the appendix A.

3.2.2 Joint Estimation and Model Specifications

In our most conservative cross-sectional specification, we address the strategic substitutability between AEM and REM (Zang, 2012; Cohen et al., 2010; Cohen et al., 2008) by cross-controlling for the alternative manipulation proxy. The two structural equations are:

$$|DA\ KM|_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \beta_2 Total\ REM_{i,t} \quad + \gamma' Controls_{i,t} + YEAR\ FE\ + SUBSEC\ FE + \varepsilon_{i,t} \quad (7)$$

$$Total\ REM_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \beta_2 |DA\ KM|_{i,t} + \gamma' Controls_{i,t} + YEAR\ FE\ + SUBSEC\ FE + \varepsilon_{i,t} \quad (8)$$

The sign on the cross-control is itself informative: a positive coefficient indicates complementarity (joint use of both channels) and a negative coefficient indicates substitution (managers trading one channel against the other), foreshadowing the dynamic analysis around the NFRD shock in Section 3.3. All standard errors are two-way clustered at the firm and year levels following Petersen (2009) and Gow et al. (2010), which ensures that significance levels (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$) are robust to both cross-sectional and time-series dependence in the residuals.

3.3 Causal Identification: NFRD Quasi-Natural Experiment and 2SLS Estimation

Section 3.2 estimates the conditional CSR-EM correlation but cannot rule out reverse causality (firms with cleaner books may invest in ESG to signal discipline) or omitted firm characteristics

(managerial ethics, stakeholder orientation) that simultaneously elevate ESG and constrain manipulation. To address these endogeneity concerns, we adopt a two-stage instrumental variables (2SLS) framework that exploits the heterogeneous bite of the NFRD across firms with different pre-existing CSR levels. The shift from voluntary to mandatory CSR disclosure that the NFRD effected in 2017 fundamentally changes the informational role of ESG reporting (Christensen et al., 2021), and the empirical strategy here tests whether this regulatory pressure produces a disciplining effect across both channels (H2.a) or instead a channel-substitution effect (H2.b).

3.3.1 Institutional Background and Treatment Timing

The EU Non-Financial Reporting Directive (Directive 2014/95/EU) was adopted in October 2014 and transposed across all member states by December 2016. It requires large public-interest entities exceeding 500 employees to disclose non-financial information on environmental, social, employee, human-rights, and anti-bribery matters in their annual management reports. Although adopted in 2014, the substantive disclosure requirements apply to financial years beginning on or after 1 January 2017. Three features make the NFRD attractive for causal identification: (i) it was enacted at the supranational level and is therefore largely exogenous to any individual firm's reporting incentives; (ii) the timing of the treatment is sharp and well-defined (post-period 2017-2025); and (iii) while the directive's legal scope is uniform across large EU firms, its substantive bite is sharply heterogeneous. As per the previous research, firms with weak pre-existing CSR engagement face material adjustment costs while firms with strong voluntary disclosure already in place do not (Fiechter et al., 2022; Grewal et al., 2019). This heterogeneity in treatment intensity provides the cross-sectional variation that drives identification.

3.3.2 Pre-Treatment Classification of Firm Exposure

Identification rests on a binary indicator of NFRD exposure constructed using only pre-treatment information. Our classification procedure rests on two design choices that depart from a naive whole-sample comparison: (i) we anchor exposure to a single pre-NFRD year rather than to a multi-year average, and (ii) we benchmark each firm's ESG performance against its own industry rather than against the full sample.

3.3.2.1 Single-Year Pre-NFRD ESG Anchor

We assign each firm a pre-NFRD ESG anchor equal to its 2016 ESG score whenever available. This single-year anchor replaces the multi-year average used in earlier drafts. Two considerations motivated the change. First, the anchor is intended to capture each firm's ESG positioning immediately prior to the policy shock. A six-year average obscures rather than characterises end-of-pre-period standing. Second, panel coverage is uneven, so a multi-year average is computed from different sample sizes for different firms, producing a sorting variable of mechanically heterogeneous precision. A single-year anchor produces a perfectly balanced one-observation-per-firm classifier.

3.3.2.2 Industry-Adjusted ESG

A second concern is that ESG scores vary systematically across industries: sectors facing acute environmental or social scrutiny disclose more, mechanically boosting their measured scores. A whole-sample split into high and low pre-NFRD ESG therefore risks confounding industry composition with treatment intensity. We use an industry-adjusted score equal to the firm's anchor minus the NAICS-subsector mean of the anchor instead:

$$AdjES\underline{G}_{i,pre} = ES\underline{G}_{i,pre} - Mean_{j}\epsilon_{s(i)}(ES\underline{G}_{j,pre}) \text{ (9)}$$

AdjES$\overline{G}$ expresses each firm's pre-NFRD ESG performance as a within-industry deviation. The exposure and post indicators are then defined as

$$Exposure_i = 1\{AdjES\underline{G}_{i,pre} \leq Q_{33}(AdjES\underline{G}_{pre})\}, Post_t = 1\{t \geq 2017\} \text{ (10)-(11)}$$

The bottom tercile of the industry-adjusted distribution defines the treated group (firms whose CSR was weak relative to their own industry in the year immediately before the NFRD, and which therefore had the greatest distance to the new disclosure standards). By anchoring classification on pre-2017 information and benchmarking each firm against its own industry, we ensure that group assignment is plausibly orthogonal to post-treatment shocks and is not confounded by cross-industry differences in baseline disclosure norms. To enter the 2SLS sample, firms must (i) have at least one valid pre-NFRD ESG observation, (ii) have non-missing earnings management measures and full controls in the firm-year, and (iii) be present in both the pre- and post-NFRD windows.

3.3.3 Stage 1: First-Stage Regression

The first stage projects contemporaneous ESG on the interaction between NFRD timing and pre-treatment exposure, together with the firm-level control vector and a high-dimensional fixed-effects structure:

$$ESG_{i,t} = \pi_0 + \pi_1(Post_t \times Exposure_i) + \gamma' X_{i,t} + \alpha_i + \lambda_t + \upsilon_{i,t} \quad \text{(12)}$$

where $X_{i,t}$ is the firm-level control vector (Size, MB, ROA, Lev, CFO_scaled, Log_Age, Analyst_Cov, Big4); α_i denotes firm fixed effects, which absorb time-invariant heterogeneity and mechanically subsume the time-invariant Exposure indicator; and λ_t denotes year fixed effects, which absorb common macroeconomic shocks and subsume the year-invariant Post indicator. The coefficient of interest is π_1 on the Post × Exposure interaction, capturing the differential ESG change among treated firms after NFRD implementation. Standard errors are two-way clustered at the firm and year levels. For sensitivity, we also report a specification that replaces firm and year fixed effects with NAICS subsector fixed effects, allowing the main effects of Post and Exposure to be identified separately.

Stage 1 has two purposes. First, it provides a first-stage strength test: we report the F-statistic on the excluded instrument and assess it against the Stock and Yogo (2005) critical values (F > 10 for relevance, F > 23 for the 5% maximum-bias criterion), together with the

heteroskedasticity-robust Kleibergen-Paap rk Wald F. Second, it generates the fitted values $\widehat{ESG}_{i,t}$, which represent the component of contemporaneous ESG mechanically driven by the NFRD shock, purged of endogenous variation. Our prior expectation is that π_1 is large and positive: low-CSR firms had the greatest scope to improve under mandatory disclosure pressure, and the literature documents systematic catch-up by laggards in the post-NFRD period (Fiechter et al., 2022; Grewal et al., 2019).

3.3.4 Stage 2: Second-Stage Regression

The second stage substitutes the fitted values from Stage 1 in place of the endogenous ESG regressor and estimates the structural CSR-EM relationship:

$$EM_{i,t} = \theta_0 + \theta_1 \widehat{ESG}_{i,t} + \gamma' X_{i,t} + \alpha_i + \lambda_t + \varepsilon_{i,t} \quad (13)$$

where $EM_{i,t}$ is either $|DA\ KM|_{i,t}$or $Total\ REM_{i,t}$. The structural parameter $\theta_1$is the local average treatment effect (LATE) on the compliers (firms whose CSR was actually moved by the NFRD). The ESG term entering the right-hand side of equation (13) denotes the fitted values from Stage 1 rather than the raw endogenous ESG variable, in the standard 2SLS sense. The two stages are estimated jointly, since a manual two-stage procedure produces correct point estimates but incorrect standard errors that fail to reflect the additional uncertainty introduced by first-stage prediction.

3.3.5 The Three Conditions for a Valid Instrument

IV validity rests on three classical requirements (Wooldridge, 2002; Roberts et al., 2013): relevance, exclusion, and exogeneity. We discuss each in turn.

3.3.5.1 Relevance

Relevance requires $\pi_1 \neq 0$ in equation (12), meaning that the instrument must move the endogenous regressor after netting out controls and fixed effects. This is the only IV requirement directly testable from the data; we test it via the first-stage F-statistic, treating F > 10 as the conventional rule of thumb (Stock and Yogo, 2005) and F > 23 as the 5% maximum-bias threshold. We additionally report the Kleibergen-Paap rk Wald F. Substantively, relevance is well-supported: the NFRD imposes binding disclosure obligations most costly to firms with weak pre-existing CSR practices, and both Fiechter et al. (2022) and Grewal et al. (2019) document substantial post-NFRD increases in ESG concentrated among less-prepared firms.

3.3.5.2 Exclusion Restriction

Exclusion requires that the instrument affect EM only through ESG. The principal threat is that the NFRD also improved firms' broader information environment, such as internal controls, external scrutiny, mandatory risk-management disclosure etc., in ways that could mechanically constrain manipulation independently of substantive CSR change. We address this in three ways. First, the analyst-coverage and Big4 controls in $X_{i,t}$ partially control for monitoring

intensity. Second, firm fixed effects absorb time-invariant differences in baseline disclosure infrastructure. Third, we re-estimate the 2SLS, dropping the Big4 control to isolate the structural coefficient from any compositional dependence on the auditor channel (Section 3.3.7.4). Notwithstanding these mitigations, exclusion is fundamentally untestable in a just-identified model. Following Larcker and Rusticus (2010), we treat the IV results as a complement to rather than a replacement for the OLS baseline, and interpret convergence between OLS and 2SLS as indirect support for the structural specification.

3.3.5.3 Exogeneity

Exogeneity requires that the instrument be uncorrelated with all unobserved determinants of EM. The Post component is mechanically exogenous as a calendar indicator determined by the supranational EU policy process. The Exposure component is determined entirely from pre-NFRD ESG data (with the anchor benchmarked within the firm's own industry), so it cannot inherit post-treatment variation in EM. The interaction Post × Exposure inherits exogeneity from its constituent parts, conditional on firm and year fixed effects. To formally verify the appropriateness of the IV approach, we report the Wu-Hausman test (Hausman, 1978; Wu, 1973), which tests the null that ESG is exogenous in the structural equation; rejection indicates 2SLS is required. We treat the result as complementary rather than as a binding decision rule, since failure to reject can reflect insufficient power rather than genuine exogeneity.

3.3.6 Supporting Analysis: Reduced-Form Difference-in-Differences

As a transparent intermediate result that remains valid under weak-instrument concerns, we report the reduced-form regression in which the outcome is regressed directly on the instrument:

$$EM_{i,t} = \delta_0 + \delta_1(Post_t \times Exposure_i) + \gamma' X_{i,t} + \alpha_i + \lambda_t + \varepsilon_{i,t} \quad (14)$$

δ_1 is the canonical difference-in-differences (intent-to-treat) estimator, capturing the average differential change in EM among low-CSR firms regardless of whether they actually improved their CSR. It is mechanically related to the 2SLS coefficient by the Wald identity $\theta_1 = \delta_1/\pi_1$. The reduced form serves three purposes: it provides a model-free check that the NFRD shock had a detectable EM effect, it is robust to weak-instrument bias, and it provides a sanity check on the joint estimation. Under H1.a (transparent reporting), increased CSR constrains opportunistic reporting and $\delta_1$is negative; under H1.b (moral licensing), CSR functions as a reputational shield and $\delta_1$is positive; under H2.b (substitution), the predictions diverge across channels, with AEM falling and REM rising.

3.3.7 Robustness and Validity Tests

We report seven complementary robustness checks. Each is summarised in one paragraph. Full coefficient tables are reported with the corresponding results in Section 5.

3.3.7.1 Event-Study and Pre-Trends

Although 2SLS does not require parallel pre-trends in the strict DiD sense, the reduced-form complement does, and the absence of differential pre-trends provides indirect support for instrument exogeneity. We estimate a dynamic event-study that replaces the binary Post × Exposure with year-specific interactions $D_{t,k} \times Exposure_i$, with k = −1 omitted as the reference. Coefficients on negative leads ($k \leq -2$) should be statistically indistinguishable from zero under parallel trends. Event-time observations beyond ±5 years are binned to prevent coefficient instability from sparse tails.

3.3.7.2 Placebo Tests

As a falsification test, we re-estimate the reduced form on the pre-NFRD subsample (2010-2016) only, designating 2013 as a placebo treatment date. A statistically insignificant coefficient on Placebo Post × Exposure is consistent with the absence of pre-existing differential dynamics that could spuriously generate the headline result. The placebo first stage on ESG is also estimated to confirm that no exogenous CSR shock is detectable at the placebo date.

3.3.7.3 Alternative Treatment Cutoffs

The bottom-tercile threshold is consistent with the prior literature on heterogeneous regulatory effects (Fiechter et al., 2022), but it is fundamentally arbitrary. We therefore re-estimate both stages with two alternative thresholds (bottom quartile and bottom half), in each case applied to the industry-adjusted pre-NFRD ESG distribution. We additionally replace the binary indicator with a continuous treatment-intensity measure, $Intensity_i$= 1 - rank($AdjES\bar{G}_i$,pre)/N, ranging from 0 (very-high pre-NFRD industry-adjusted ESG) to 1 (very-low). The continuous instrument is the interaction $Intens\ Post_{i,t} = Intensity_i \times Post_t$.

3.3.7.4 Robustness to Auditor-Data Missingness

The LSEG Auditor field is unpopulated for approximately three-quarters of cleaned firm-years, so any specification including Big4 is estimated on a meaningfully smaller sample. We re-estimate the main 2SLS dropping Big4 from the control vector to recover the larger sample and isolate the structural coefficient from any compositional dependence on auditor monitoring. Stability of the structural coefficient across the two specifications would indicate that the auditor-data limitation does not threaten the headline inference.

3.3.7.5 Industry-Cell Minimum-Size Sensitivity

The baseline implementation of the Kothari and Roychowdhury models requires at least ten observations per NAICS subsector-year cell. To verify that this threshold does not produce noisy benchmarks for thinly populated industries, we re-estimate the models with stricter thresholds of 15 and 20 observations per cell and confirm that the noise content of the |DA_KM|

proxy declines as the threshold tightens (Appendix F, Appendix G), and that the headline H1 conclusion is unaffected because the noise component is additive to the residual rather than correlated with the regressor of interest.

3.3.7.6 Propensity-Score Reweighted 2SLS

Firm and year fixed effects absorb time-invariant differences between treated and control firms but cannot mechanically eliminate differential pre-treatment trajectories on observable characteristics. To address the imbalance documented in the pre-treatment balance test, we estimate a logit propensity-score model $\Pr(\text{Exposure}_i = 1) = \Lambda(\gamma_0 + \gamma' \text{Pre\_X}_i)$, where Pre_X_i is the firm-level pre-NFRD average of the eight controls (Size, MB, Lev, ROA, CFO_scaled, Log_Age, Analyst_Cov, Big4). Inverse-probability-of-treatment weights are $1/\hat{p}_i$ for treated and $1/(1 - \hat{p}_i)$ for control firms, trimmed at the 1st and 99th percentiles. We then re-estimate the joint 2SLS specification with IPTW weights. Stability of the structural coefficient indicates that the pre-treatment imbalance is not driving the result.

3.3.7.7 Within-Industry Identification

The bottom-tercile classification implicitly imposes a homogeneous-treatment-effect assumption, which may dilute effects concentrated in the very lowest pre-CSR firms. Fiechter et al. (2022) establish that the real effects of the NFRD on CSR activities are concentrated in firms with low pre-directive CSR. A related concern is that the structural coefficient could, in principle, be identified off cross-industry rather than within-industry variation in treatment intensity, which would confound the within-industry causal effect with industry-level heterogeneity in earnings management behaviour. To address this, we re-estimate the 2SLS on the subsample of NAICS subsectors that contain both treated and control firms, forcing identification to operate within industries that contain both groups. Stability of the structural coefficient under this restriction rules out the interpretation that the headline result is driven by cross-industry confounding.

3.3.8 Acknowledged Limitations of the Identification Strategy

Several limitations of the 2SLS approach merit explicit acknowledgement. First, the absence of a clean external control group (such as comparable non-EU firms unaffected by the NFRD) means identification rests entirely on within-EU heterogeneity in treatment intensity. Most European studies use US firms as a control group when studying NFRD effects, but our European-only research design precludes this approach. Our design therefore compares more-exposed and less-exposed EU firms (both subject to the directive's legal scope) with exposure defined relative to industry peers in the year immediately preceding the directive. The resulting estimates capture the differential impact of more aggressive compliance rather than the absolute effect of the regulation per se. Second, the just-identified nature of our 2SLS specification precludes overidentification tests (Sargan, Hansen J) that could provide additional statistical evidence on the exclusion restriction. The case for exclusion therefore rests on institutional argumentation rather than statistical testing. Third, the structural coefficient θ_1 identifies the local average treatment effect on the compliers rather than the average effect

across all firms. Extrapolation requires additional assumptions about treatment-effect heterogeneity. Finally, the European post-treatment window (2017-2025) is relatively short, and treatment effects may continue to evolve under the successor Corporate Sustainability Reporting Directive (CSRD). We interpret the 2SLS estimates with these caveats in mind, presenting them as causal evidence conditional on the assumptions of Section 3.3.5 and as a complement to the broader OLS analysis of Section 3.2.

4. Data and Descriptive Statistics

This Section describes the empirical foundation of the study. Section 4.1 documents the construction of the analytical samples and the attrition pathway from the raw extract. Section 4.2 reports descriptive statistics for the firm-level financial and CSR variables across the samples used in the multivariate analysis. Section 4.3 presents the industry composition of the baseline sample. Section 4.4 reports correlation patterns. Section 4.5 introduces additional descriptive evidence specific to the NFRD analytical sample used in Section 5.

4.1 Sample Construction and Attrition

Table 1 reports the sample-selection waterfall in two panels. Panel A documents the construction of the baseline regression sample, and Panel B the additional restrictions used to build the NFRD analytical sample. The initial extract from LSEG Workspace yields 97,485 firm-year observations covering listed entities in the 27 EU member states over 2010-2025. Application of the sectoral filters, such as excluding financial institutions and firms in regulated or sin industries, together with the explicit 2010-2025 fiscal-year window leaves 84,511 industrial firm-year observations. The construction of lagged scaling variables and changes in revenue further reduces the panel to 28,282 firm-years for which the Kothari and Roychowdhury model inputs can be computed. The cross-sectional estimation of normal accruals and abnormal real activities at the NAICS subsector-year level requires a minimum of ten observations per estimation cell, successfully estimating earnings management metrics for 10,030 firm-years. After winsorization at the 1st and 99th percentiles and listwise deletion on the full control vector, the final baseline regression sample consists of 16,067 firm-year observations across 25 NAICS subsectors and the 27 EU member states, spanning 1,944 distinct firms.

Panel B reports the construction of the NFRD analytical sample. To enable the difference-in-differences and instrumental-variables design around the directive's 2017 effective date, firms must satisfy three additional conditions beyond the baseline sample: (i) a usable pre-NFRD ESG anchor observation (see Section 4.5.1); (ii) presence in both the pre- and post-NFRD windows so that the within-firm treatment comparison is well-defined; and (iii) availability of the full vector of regression controls. Of the 1,944 firms in the baseline sample, 323 have a usable anchor-year ESG observation, and after imposing the balanced-panel requirement, the NFRD analytical sample consists of 223 firms (2,836 firm-year observations). The pre-treatment exposure classification, based on the bottom tercile of the industry-adjusted pre-

NFRD anchor, identifies 70 treated firms (852 firm-years) and 153 control firms (1,984 firm-years).

Table 1. Sample Selection Waterfall

Step	Firm-years	Firms
Panel A: Baseline regression sample		
1. Initial raw import	97,485	—
2. After sectoral filters (excl. financials, sin)	84,511	—
3. After lag/delta construction	28,282	—
4. Strict overlapping sample (both AEM & REM proxies estimable)	10,030	—
5. Final baseline regression sample	16,067	1,944
Panel B: NFRD analytical sample		
A. Final baseline regression sample	16,067	1,944
B. Firms with usable pre-NFRD ESG anchor	3,199	323
C. After requiring both pre- and post-NFRD presence	2,836	223
Of which: Treated (bottom-tercile industry-adj. anchor)	852	70
Of which: Control	1,984	153

Notes: Panel A reports the construction of the baseline regression sample. The Step 5 row reports the sample after winsorization at the 1st and 99th percentiles; the apparent expansion relative to Step 4 reflects the fact that EM-estimation success and final-control availability are imposed on overlapping but non-identical subsets of the panel. Panel B reports the additional restrictions used to construct the NFRD analytical sample for Section 5: a usable pre-NFRD ESG anchor (Section 4.5.1), presence in both the pre- and post-NFRD windows, and full control-vector availability.

4.2 Descriptive Statistics

Table 2 reports descriptive statistics for the principal continuous variables in two panels: Panel A is the baseline regression sample, and Panel B is the strict overlapping sample (N = 1,098 firm-years) used in the trade-off analysis of Section 5.2. The strict overlapping sample retains only firm-years with non-missing values on both earnings management proxies and on the full control vector.

In the baseline sample, the mean absolute value of discretionary accruals (|DA_KM|) is 0.252 with a median of 0.059, reflecting the right-skewed distribution typical of accrual-based earnings management proxies in European panels. The aggregate real earnings management measure (Total_REM) has a mean of -0.036 and a standard deviation of 0.330, broadly comparable to the magnitudes reported by Roychowdhury (2006) for U.S. firms and by subsequent European applications. The Roychowdhury components have a mean near zero by construction (Abn CFO 0.003, Abn PROD -0.041, Abn DISEXP -0.009).

The cross-section displays substantial variation in CSR engagement. The aggregate ESG Score has a mean of 58.97 and a standard deviation of 18.64, essentially identical to the 59.28 reported by Velte (2021) for a 568-firm EU panel, providing reassurance that the LSEG distribution is

representative of the broader European listed-firm universe. The disaggregated pillar scores reveal a familiar hierarchy: Social (mean 64.45), Environmental (58.01), and Governance (51.52). The lower mean of Governance reflects the well-documented heterogeneity of corporate-governance regimes across EU member states (Liang and Renneboog, 2017). Among firm-level controls, the average firm has a log-market-capitalization of 19.34 (roughly EUR 250 million), a market-to-book ratio of 2.38, a leverage ratio of 0.19, and an ROA of 0.007. Analyst coverage averages 5.04 analysts per firm-year. Conditional on non-missing auditor identification, 84.85% of firm-years are audited by a Big4 firm, broadly consistent with the 77.93% reported by Velte (2021) for European public-interest entities.

Panel B reports the strict overlapping sample. As expected, this stricter sample exhibits a compositional shift toward larger, better-monitored firms: mean Size rises to 22.48 (vs. 19.34 in the baseline), mean Analyst Coverage to 15.50 (vs. 5.04), and the tails of the earnings management proxies are noticeably tighter. The Big4 share remains stable at 0.84. We retain the strict overlapping sample for the trade-off analysis of Section 5.2 because the joint identification of AEM and REM coefficients requires complete data on both proxies.

Table 2. Descriptive Statistics

Variable	N	Mean	SD	Min	P25	Median	P75	Max
Panel A: Baseline sample								
|DA_KM| (AEM)	5,251	0.252	0.676	0.001	0.023	0.059	0.158	4.960
Total_REM	3,474	-0.036	0.330	-0.791	-0.244	-0.055	0.122	1.218
Abn CFO	3,474	0.003	0.087	-0.469	-0.030	0.002	0.037	0.342
Abn PROD	3,474	-0.041	0.307	-0.640	-0.237	-0.062	0.097	1.158
Abn DISEXP	3,474	-0.009	0.108	-0.225	-0.076	-0.021	0.034	0.483
ESG Score	5,599	58.97	18.64	8.30	46.94	61.54	73.46	89.85
Environmental Pillar	5,598	58.01	24.40	0.00	41.10	61.64	77.41	95.94
Social Pillar	5,598	64.45	21.01	5.43	50.73	68.30	81.43	95.64
Governance Pillar	5,599	51.52	22.37	5.39	33.53	52.56	69.70	92.09
Size (log Mkt Cap)	16,067	19.34	2.40	14.28	17.54	19.20	21.06	24.85
Market-to-Book	16,067	2.38	3.81	0.11	0.75	1.31	2.46	29.42
ROA	15,835	0.007	0.140	-1.092	-0.008	0.026	0.060	0.315
Leverage	16,067	0.194	0.149	0.000	0.074	0.169	0.289	0.709
CFO (scaled)	11,997	0.055	0.125	-0.759	0.024	0.067	0.109	0.411
log(1+Firm Age)	13,189	3.26	0.74	1.10	2.89	3.33	3.78	4.68
Analyst Coverage	16,067	5.04	7.21	0.00	0.00	2.00	7.00	29.00

Variable	N	Mean	SD	Min	P25	Median	P75	Max
Big4 Auditor	3,894	0.849	0.359	0.000	1.000	1.000	1.000	1.000
Panel B: Strict overlapping sample (trade-off analysis)								
|DA_KM| (AEM)	1,098	0.207	0.573	0.001	0.019	0.049	0.141	4.960
Total_REM	1,098	-0.058	0.320	-0.791	-0.260	-0.072	0.086	1.218
ESG Score	1,098	68.72	13.74	15.94	60.33	70.91	78.96	89.85
Size (log Mkt Cap)	1,098	22.48	1.49	18.23	21.48	22.55	23.62	24.85
Market-to-Book	1,098	2.75	3.54	0.11	0.94	1.67	2.99	29.42
ROA	1,098	0.036	0.065	-0.426	0.013	0.035	0.063	0.315
Leverage	1,098	0.210	0.125	0.000	0.116	0.203	0.289	0.669
CFO (scaled)	1,098	0.092	0.066	-0.119	0.056	0.084	0.116	0.411
log(1+Firm Age)	1,098	3.44	0.74	1.10	3.04	3.50	3.99	4.68
Analyst Coverage	1,098	15.50	7.74	0.00	8.00	16.00	22.00	29.00
Big4 Auditor	1,098	0.840	0.367	0.000	1.000	1.000	1.000	1.000

Notes: All continuous variables are winsorized at the 1st and 99th percentiles. Panel A reports the baseline regression sample (final_data); N varies across rows due to missingness on individual variables. Panel B reports the strict overlapping sample (final_subst_data) used in the trade-off analysis of Section 5.2, which by construction conditions on complete data.

4.3 Industry Composition

Table 3 reports the industry composition of the baseline regression sample at the NAICS subsector level, with mean ESG, AEM, and REM values by subsector. The sample spans 25 distinct subsectors, with the largest representation in Real Estate (1,712 firm-years), Professional Services (1,632), Chemical Manufacturing (917), Computer and Electronic Manufacturing (806), and Machinery Manufacturing (645). Substantial cross-industry variation is visible in both CSR engagement and earnings management. ESG scores range from 46.6 in Publishing Industries to 68.6 in Primary Metal Manufacturing, with higher engagement in capital-intensive and traditionally regulated industries (Utilities 65.96, Primary Metals 68.65, Electrical Equipment 65.81) than in service-oriented sectors (Professional Services 56.52, Publishing 46.63).

The cross-industry distribution of earnings management is also revealing. Several capital-intensive industries display substantially higher mean |DA_KM| than service-oriented sectors: Heavy Engineering Construction (0.986), Construction of Buildings (0.967), and Support Activities for Transportation (1.540) all exceed 0.9, compared to Computer and Electronic Manufacturing (0.074) and Machinery Manufacturing (0.095). This pattern suggests that any negative cross-sectional ESG-EM correlation observed in the raw data is likely driven by inter-industry differences rather than within-industry behavioral relationships, reinforcing the

methodological importance of the NAICS subsector fixed effects included in all multivariate specifications. The Support Activities for Transportation subsector exhibits an unusually high mean |DA_KM| (1.540), reflecting the small sample size (N = 193) and a small number of extreme observations that pass the winsorization threshold.

Table 3. Industry Composition of the Baseline Sample

Industry (NAICS Subsector)	N	Mean ESG	Mean AEM	Mean REM
Real Estate	1,712	58.50	0.240	0.018
Professional, Scientific, and Technical Services	1,632	56.52	0.103	-0.053
Chemical Manufacturing	917	62.12	0.094	-0.027
Computer and Electronic Product Manufacturing	806	57.71	0.074	-0.053
Machinery Manufacturing	645	53.54	0.095	-0.060
Transportation Equipment Manufacturing	638	60.83	0.256	-0.062
Publishing Industries	622	46.63	0.196	-0.008
Utilities	553	65.96	0.152	-0.095
(Unspecified NAICS)	478	65.00	0.468	-0.028
Food Manufacturing	431	60.22	0.093	-0.034
Heavy and Civil Engineering Construction	422	61.29	0.986	-0.032
Merchant Wholesalers, Durable Goods	349	57.23	0.354	0.039
Construction of Buildings	311	61.54	0.967	-0.069
Merchant Wholesalers, Nondurable Goods	299	55.48	0.480	0.100
Primary Metal Manufacturing	298	68.65	0.239	-0.073
Other 10 subsectors (combined)	2,978	—	—	—

Notes: The table reports the 15 NAICS subsectors with the largest representation in the baseline regression sample. The remaining 10 subsectors are aggregated in the final row to preserve readability; the full subsector listing is available on request. "(Unspecified NAICS)" denotes firm-years for which the NAICS Subsector Name field was empty in the LSEG extract.

4.4 Correlation Analysis

Table 4 reports correlation patterns among the principal variables in two panels. Panel A reports correlations on the baseline sample, with Pearson coefficients below the diagonal and Spearman coefficients above. Panel B reports the corresponding matrix on the strict overlapping sample. Three patterns warrant attention. First, the simple correlation between ESG and earnings management is negative and highly significant for both proxies (Pearson: -0.073*** with |DA_KM| and -0.093*** with Total_REM; Spearman -0.126*** and -0.107***). This raw association is broadly consistent with the prediction of the Transparent Financial Reporting Hypothesis (H1.a). The somewhat larger Spearman than Pearson coefficients suggest that the relationship is partly driven by rank-order rather than strictly linear association.

Second, ESG correlates strongly with firm size (Pearson 0.577***), indicating that a substantial portion of the negative ESG-EM correlation could plausibly reflect the well-established size-driven reduction in earnings management documented by Watts and Zimmerman (1990) under the political-cost hypothesis. This collinearity foreshadows the importance of controlling for size and other firm characteristics in the multivariate analysis in Section 5. Third, the correlation between AEM and REM is positive but small (Pearson 0.082***; Spearman 0.088***), indicating that the two manipulation channels are not strict substitutes in the cross-section. This is consistent with the pattern documented in the international literature where the substitution dynamics emphasized by Zang (2012) and Cohen and Zarowin (2010) tend to be more pronounced in time-series settings around regulatory shocks than in pure cross-sections.

Panel B reveals that the patterns identified in the baseline sample are qualitatively preserved in the strict overlapping sample, with somewhat larger magnitudes for several pairs: the Pearson correlation between AEM and REM rises to 0.099***, and the negative ESG-AEM Pearson correlation strengthens to -0.136***. The ESG-REM correlation, by contrast, becomes statistically insignificant in the strict overlapping sample (Pearson -0.023, ns), an early hint that the negative cross-sectional REM association observed in the baseline is partly mechanical rather than behavioural.

Table 4. Correlation Matrix

Variable	|DA_KM|	Total_REM	ESG	Size	MB	ROA	Lev
Panel A: Baseline sample (Pearson below diagonal, Spearman above)							
|DA_KM| (AEM)	1	0.088***	-0.126***	-0.197***	-0.134***	-0.138***	-0.064***
Total_REM	0.082***	1	-0.107***	-0.201***	-0.179***	-0.106***	-0.087***
ESG Score	-0.073***	-0.093***	1	0.576***	-0.015	0.075***	0.099***
Size	-0.118***	-0.184***	0.577***	1	0.287***	0.266***	0.072***
MB	-0.063***	-0.134***	-0.039	0.203***	1	0.403***	-0.013
ROA	-0.034	-0.079***	0.065***	0.223***	0.263***	1	-0.062***
Lev	-0.082***	-0.093***	0.083***	0.079***	0.059**	-0.028	1
Panel B: Strict overlapping sample (Pearson below diagonal, Spearman above)							
|DA_KM| (AEM)	1	0.071**	-0.143***	-0.179***	-0.125***	-0.123***	-0.060**
Total_REM	0.099***	1	-0.046	-0.200***	-0.176***	-0.125***	-0.067**

Variable	|DA_KM|	Total_REM	ESG	Size	MB	ROA	Lev
ESG Score	-0.136***	-0.023	1	0.507***	-0.082***	0.004	0.119***
Size	-0.160***	-0.173***	0.517***	1	0.320***	0.238***	-0.016
MB	-0.059*	-0.155***	-0.123***	0.206***	1	0.438***	-0.052*
ROA	-0.054*	-0.112***	-0.007	0.219***	0.359***	1	-0.133***
Lev	-0.079***	-0.090***	0.100***	-0.004	0.006	-0.129***	1

*Notes: All correlations computed after winsorization at the 1st and 99th percentiles. Pearson product-moment correlations are reported below the diagonal; Spearman rank correlations above. Diagonal entries are 1 by construction. Stars denote two-tailed significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.10$.*

4.5 The NFRD Analytical Sample

This section reports additional descriptive evidence specific to the NFRD analytical sample used for causal identification in Section 5. Its construction is summarised in Panel B of Table 1.

4.5.1 Pre-Treatment ESG Anchor and Industry Adjustment

First, since averaging six years (2010-2016) gives an unbalanced picture (firms with longer pre-NFRD coverage are anchored on a different object than firms with only two pre-NFRD years), and does not represent the firm's actual standing immediately before the directive's effective date. We instead take each firm's 2016 ESG score as the anchor. The resulting anchor is balanced one-observation-per-firm and reflects the picture of each firm immediately prior to treatment. As Table 5b shows, 380 firms in the broader LSEG extract are anchored in 2016. Of these, 323 are also present in the baseline regression sample reported in Panel A of Table 1, and form the population from which the 223-firm NFRD analytical sample is derived after imposing the balanced-panel requirement.

Second, we adjust the anchor for systematic industry differences in CSR disclosure before sorting firms into treatment groups. ESG levels vary mechanically across industries because firms in scrutiny-heavy sectors (utilities, primary metals) disclose more and score higher regardless of underlying CSR practices. A whole-sample tercile split therefore mixes industry composition with treatment intensity. We compute Adj_ESG as the anchor-year ESG minus the NAICS-subsector mean of the anchor-year ESG, and define Exposure as the bottom-tercile indicator on this industry-adjusted score. This identifies firms that are weak CSR performers relative to their own industry peers, regardless of absolute level.

Table 5a reports the distribution of the anchor-year ESG and of the industry-adjusted anchor (Adj_ESG) across the 380 firms with a usable anchor. The anchor ESG has mean 57.98 and standard deviation 19.39, while Adj_ESG has mean approximately zero by construction and standard deviation 17.27. The bottom-tercile cutoff falls at -5.244 ESG points below the firm's industry mean. Figure 1 plots the histogram of Adj_ESG with the cutoff line marked, confirming visually that the cutoff isolates a substantively distinct cluster of within-industry CSR underperformers rather than slicing through the middle of the distribution.

Table 5a. Pre-NFRD ESG Distribution - Anchor and Industry-Adjusted

Statistic	Anchor ESG	Adj_ESG
N firms	380	380
Mean	57.98	0.000
SD	19.39	17.27
Min	—	-46.41
P10	—	-24.23
P25	—	-10.62
P33 (cutoff)	—	-5.24
Median	61.01	1.66
P75	—	12.48
P90	—	21.71
Max	—	38.09

Notes: The anchor ESG is the firm's 2016 ESG score, with fallback to the most recent available pre-2016 year if 2016 is missing. Adj_ESG is the anchor ESG minus the NAICS-subsector mean of the anchor ESG; by construction, it has a mean of approximately zero across the 380 firms. The bottom-tercile (P33) cutoff is -5.24 ESG points below the firm's industry mean. Firms below the cutoff are classified as Treated (Exposure = 1).

Table 5b. Pre-NFRD Anchor-Year Composition

Anchor year	N firms	Percent
2016	380	99.7%

Notes: Anchor-year distribution across the 380 firms with usable pre-NFRD ESG.

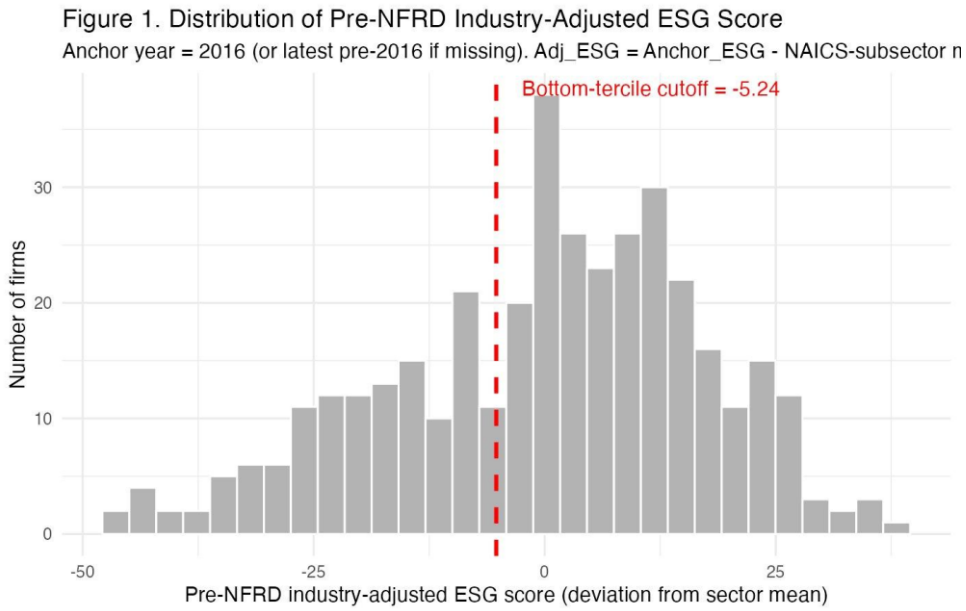


Figure 1. Distribution of Pre-NFRD Industry-Adjusted ESG Score (Adj_ESG) across the 380 firms with a usable anchor. The dashed vertical line marks the bottom-tercile cutoff (-5.24). Firms left of the line are classified as Treated.

4.5.2 Pre-Treatment Balance

Table 6 reports the formal pre-treatment balance test: two-sample Welch t-tests on the mean of each variable across treated and control firms in the pre-NFRD window only. As expected, all four ESG-related variables display large and statistically significant differences across groups (t-statistics ranging from 12.76 to 29.37, all $p < 0.001$). These are mechanical consequences of the sorting rule and do not threaten identification. The two earnings management proxies are statistically indistinguishable in the pre-treatment period (|DA_KM| difference +0.061, $p = 0.251$; Total_REM difference +0.032, $p = 0.490$), supporting the basic DiD design.

The test also reveals that treated and control firms differ on several non-ESG observables in the pre-treatment period: treated firms are smaller (Size difference -0.94, $p < 0.001$), have less analyst coverage (-6.10, $p < 0.001$), are marginally less likely to be audited by Big4 (0.76 vs 0.86, $p = 0.050$), have somewhat higher market-to-book ratios (+0.53, $p = 0.002$), have higher scaled CFO (+0.012, $p = 0.010$), and are marginally younger (log_Age 3.28 vs 3.43, $p = 0.016$). On the other hand, they are statistically indistinguishable on leverage and ROA. The pre-treatment imbalance on size, monitoring, and audit quality motivates two design choices that we maintain throughout Section 5: inclusion of these variables as controls in every multivariate specification, and a propensity-score-weighted (IPTW) robustness analysis that reweights the sample on pre-treatment observables. The 2SLS estimates from the rebalanced sample are quantitatively similar to the main estimates, providing reassurance that the imbalance is not driving the structural results.

Table 6. Pre-Treatment Balance Test

Variable	Mean (Control)	Mean (Treated)	Difference	t-stat	p-value
|DA_KM| (AEM)	0.127	0.188	+0.061	-1.153	0.251
Total_REM	-0.094	-0.062	+0.032	-0.693	0.490
ESG Score	66.07	37.69	-28.38	29.37	<0.001
Environmental Pillar	69.37	37.56	-31.81	21.02	<0.001
Social Pillar	69.44	38.40	-31.04	24.63	<0.001
Governance Pillar	56.28	37.90	-18.38	12.76	<0.001
Size (log Mkt Cap)	22.49	21.55	-0.94	10.21	<0.001
Market-to-Book	1.86	2.39	+0.53	-3.11	0.002
ROA	0.032	0.026	-0.006	1.05	0.295
Leverage	0.214	0.223	+0.008	-0.82	0.414
CFO (scaled)	0.078	0.090	+0.012	-2.58	0.010
log(1+Firm Age)	3.43	3.28	-0.148	2.43	0.016
Analyst Coverage	19.56	13.46	-6.10	11.14	<0.001
Big4 Auditor	0.864	0.757	-0.107	1.99	0.050

Notes: Two-sample Welch t-tests on pre-NFRD observations (Actual_Year < 2017) within the NFRD analytical sample. Treated firms are those below the bottom-tercile cutoff of the industry-adjusted pre-NFRD ESG anchor (Adj_ESG ≤ −5.24); Control firms are the remainder. The statistically significant pre-treatment imbalance on size, market-to-book, scaled CFO, log_Age, analyst coverage, and Big4 motivates the IPTW robustness analysis reported in Table 10 of Section 5.

4.5.3 Sample Composition by Year

Table 7 reports the year-by-year composition of the NFRD analytical sample by treatment group. Both groups maintain stable representation across the sample window: treated firm-years rise from 3 in 2011 to a plateau of 62-67 from 2014 onward, while control firm-years rise from 11 in 2011 to a stable 144-148 from 2014. The relative stability of group composition from 2014 onward reflects the natural growth of LSEG ESG coverage rather than systematic firm entry or exit. This stable composition is important for the validity of the DiD design: a sharp shift in sample composition coincident with the 2017 treatment date would confound the treatment effect with composition-induced variation, and we observe no such shift.

The industry composition of the NFRD sample by treatment group (full table available on request) confirms the expected pattern: treated firms are spread across both service-oriented and capital-intensive industries, with the largest treated concentrations in Chemical Manufacturing, Professional Services, and Real Estate, while several subsectors are dominated entirely by control firms (Petroleum and Coal Products, Broadcasting and Content Providers, Electrical Equipment) and a few smaller subsectors are dominated by treated firms (Paper Manufacturing, Oil and Gas Extraction, Fabricated Metal). Subsectors entirely composed of either treated or control firms are absorbed by NAICS subsector fixed effects in the industry-FE specifications and by firm fixed effects in the preferred firm-FE specifications, and a robustness check restricting the sample to mixed-treatment industries is reported in Section 5.

Table 7. NFRD Sample Composition by Year

Year	N (Control)	N (Treated)	Total
2011	11	3	14
2012	121	47	168
2013	126	47	173
2014	143	56	199
2015	145	64	209
2016	147	62	209
2017	144	62	206
2018	144	64	208
2019	145	66	211
2020	145	64	209
2021	148	65	213
2022	147	67	214
2023	147	66	213
2024	147	65	212
2025	124	54	178

Notes: Year-by-year composition of the NFRD analytical sample by treatment group. N denotes firm-year observations. Stable group composition through the 2017 break date is important for the validity of the DiD design.

4.5.4 Raw Time Trends in ESG and Earnings Management

Figures 2-4 plot the raw mean trajectories of ESG, |DA_KM| (AEM), and Total_REM by treatment group across the sample window. Figure 2 shows that both groups exhibit gradual improvements in mean ESG over time, but the treated group exhibits a markedly steeper trajectory beginning around 2017, consistent with the NFRD's compliance bite. By the end of the sample window, the treated group's mean ESG has risen from approximately 26 in 2011 to approximately 65 in 2025, while the control group's mean rises more gradually from 63 to 74. The post-2017 acceleration in the treated group's ESG trajectory is the central institutional pattern that the first-stage regression in Section 5 quantifies formally.

Figures 3 and 4 plot the analogous time trends for the two earnings management proxies. Both proxies exhibit substantially more year-to-year volatility than ESG, reflecting the inherent noisiness of accruals and operational manipulation measures. The visual evidence does not show a clear differential break in either earnings management measure at the 2017 treatment date, foreshadowing the structural results presented in Section 5. Figure 3 displays a large outlier point in the treated group's mean |DA_KM| for 2012, reflecting the small treated sample in that year (N = 47) and the presence of a small number of extreme observations that pass the winsorization threshold. The REM time series in Figure 4 reveals that the two groups' means roughly track each other throughout the window, with a brief divergence around the treatment date that subsequently closes.

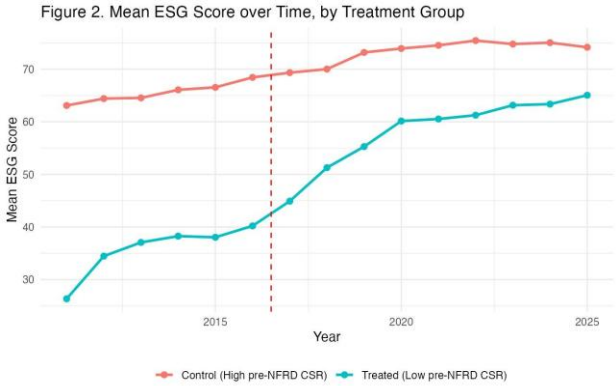


Figure 2. Mean ESG Score by year for the treated group (bottom-tercile industry-adjusted pre-NFRD ESG) and control group within the NFRD analytical sample. Vertical dashed line marks the NFRD effective date (2017). The treated group exhibits a markedly steeper post-2017 ESG trajectory than the control group, providing visual support for the first-stage relationship formalised in Section 5.

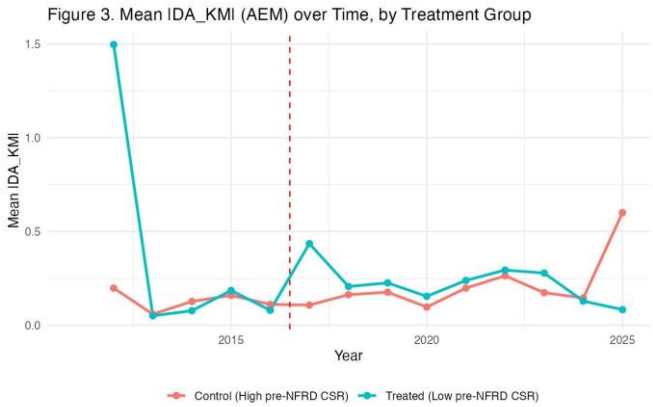


Figure 3. Mean absolute discretionary accruals (|DA_KM|, the AEM proxy) by year for the treated and control groups within the NFRD analytical sample. Vertical dashed line marks the NFRD effective date (2017). The series displays substantial year-to-year volatility, reflecting the noisy nature of discretionary-accruals estimates at the group-mean level.

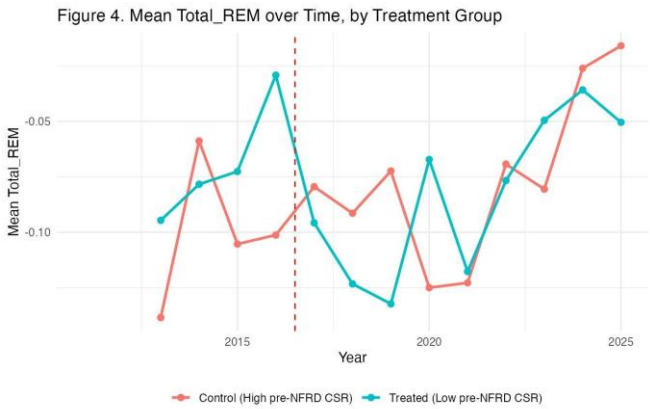


Figure 4. Mean Total_REM by year for the treated and control groups within the NFRD analytical sample. Vertical dashed line marks the NFRD effective date (2017). The two groups roughly track each other across the window, with a brief divergence around the treatment date that subsequently closes.

5. Results

This section reports the multivariate evidence on the relationship between CSR performance and earnings management in the European listed-firm panel described in Section 4. The empirical analysis proceeds in four stages, each mapped onto one of the four hypotheses developed in Section 2.3. Section 5.1 reports the multivariate OLS evidence on the baseline cross-sectional association between aggregate ESG and the two earnings management proxies, testing the competing predictions of the Transparent Financial Reporting Hypothesis (H1.a) against the Opportunistic Financial Reporting Hypothesis (H1.b). Section 5.2 introduces the joint-substitution specifications that hold the alternative manipulation channel constant, sharpening the H1 test by isolating the channel-specific CSR effect. Section 5.3 turns to the causal evidence from the 2017 NFRD shock, using the two-stage least-squares framework to test the disciplining prediction of H2.a and the substitution prediction of H2.b. Section 5.4 reports the six robustness specifications that probe the sensitivity of the headline causal estimate. Section 5.5 synthesises the evidence against each of the four hypotheses and discusses the most plausible reasons for the predominantly null findings, anchored in the prior studies reviewed in Section 2.

5.1. The Baseline Association Between CSR and Earnings Management (H1.a vs H1.b)

The first set of multivariate regressions tests the competing predictions of the Transparent Financial Reporting Hypothesis (H1.a) and the Opportunistic Financial Reporting Hypothesis (H1.b) on the conditional cross-sectional association between aggregate ESG and the two earnings management channels. The baseline H1 specification is an OLS regression of the earnings management proxy on contemporaneous aggregate ESG and the full firm-level control vector defined in Section 3, namely Size (log market capitalisation), Market-to-Book, ROA, Leverage, scaled CFO, log(1+Firm Age), Analyst Coverage, and a Big4 auditor indicator, together with NAICS subsector fixed effects and year fixed effects. Standard errors are two-way clustered at the firm and year levels. Crucially, this baseline regression does not include

the alternative earnings management proxy on the right-hand side. The cross-channel control is introduced separately in Section 5.2, where the joint-substitution specification tests whether holding the alternative manipulation channel constant alters the H1 inference. The baseline of Section 5.1 is therefore the standard unconditional H1 regression. Two parallel regressions are estimated: column (1) of Table 8 takes the absolute value of performance-matched discretionary accruals (|DA_KM|), constructed following the cross-sectional Kothari et al. (2005) model, as the dependent variable, and column (2) takes total real earnings management (Total_REM), constructed following Roychowdhury (2006), as the dependent variable. This dual-channel design is essential for the H1 test, since the moral track described by Kim et al. (2012) and Jones (1995) predicts that CSR commitment constrains opportunism along both channels, not only one. Effective samples comprise 1,545 firm-years for the AEM regression and 1,413 firm-years for the REM regression, drawn from the 16,067 firm-year baseline panel after listwise deletion on each proxy and the full control vector.

Table 8. Aggregate ESG and Earnings Management (Main H1)

	Dependent variable: |DA_KM| (AEM) / Total REM	
Variable	**AEM**	**REM**
ESG Score	-0.000	0.000
	(0.001)	(0.001)
Size (log Mkt Cap)	-0.028	-0.013
	(0.025)	(0.019)
Market-to-Book	0.007*	-0.001
	(0.004)	(0.004)
ROA	-0.233	-0.139
	(0.187)	(0.127)
Leverage	-0.490***	-0.295**
	(0.168)	(0.128)
CFO (scaled)	-0.029	-1.313***
	(0.269)	(0.350)
log(1+Firm Age)	-0.035	-0.005
	(0.023)	(0.022)
Analyst Coverage	-0.003	-0.004
	(0.003)	(0.003)
Big4 Auditor	-0.042	-0.022

	(0.042)	(0.050)
Year FE	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes
Observations	1,545	1,413
R-squared	0.269	0.143
Adj. R-squared	0.245	0.115

*Notes: OLS estimates of the conditional association between aggregate ESG performance and earnings management. Column (1) regresses absolute performance-matched discretionary accruals on ESG and the firm-level control vector; column (2) does the same with the aggregate real-earnings-management proxy as the dependent variable. Both specifications include year and NAICS subsector fixed effects (coefficients omitted for brevity). Note: Robust standard errors clustered at the firm and year levels in parentheses. *** p<0.01, ** p<0.05, * p<0.10.*

The headline finding from the baseline OLS specification is that the estimated coefficient on aggregate ESG is statistically indistinguishable from zero in both regressions. The point estimate is essentially zero in the AEM regression (standard error of approximately 0.001) and equally negligible in the REM regression, with neither approaching conventional significance thresholds. The bivariate correlations reported in Table 4 of Section 4 (negative and significant for both proxies, namely Pearson r equal to minus 0.073 for AEM and minus 0.093 for REM) therefore do not survive the introduction of the firm-level control vector and the saturated fixed-effects structure. Read jointly, the baseline OLS results provide no support for the directional prediction of the Transparent Financial Reporting Hypothesis (H1.a) that ESG commitment is associated with a lower magnitude of earnings management across both channels, and equally no support for the Moral Licensing prediction of the Opportunistic Financial Reporting Hypothesis (H1.b) that ESG engagement is associated with a higher magnitude of manipulation. The European cross-sectional evidence is at odds with the negative association documented for U.S. firms by Kim et al. (2012) but aligns with the contextual contingency emphasised by Chih et al. (2008), who show that the CSR and earnings management relationship is highly dependent on the institutional environment, and with Liang and Renneboog (2017), who demonstrate that civil-law jurisdictions exhibit a higher baseline of compliance-driven CSR engagement that may weaken the moral-track mechanism that operates in common-law settings.

Although the variable of interest is null, the firm-level controls in the baseline OLS specification load in directions consistent with prior earnings management research. Leverage

carries a negative and highly significant coefficient on AEM and a smaller but still significant negative coefficient on REM, which is consistent with the discipline imposed by external creditors on managerial reporting discretion. Cash flow from operations enters strongly negative in the REM regression, reflecting the construction of the abnormal cash-flow component of the Roychowdhury proxy. The market-to-book ratio loads marginally positive on AEM, consistent with the well-known association between high-growth firms and larger absolute accruals documented in the methodological discussion of Kothari et al. (2005). Firm size, return on assets, firm age, analyst coverage, and the Big4 indicator do not load significantly in either regression, although their signs align with the patterns reported in earlier work. The model fit, as indicated by the adjusted R-squared values of 0.245 in the AEM regression and 0.115 in the REM regression, is relatively low, suggesting that a significant portion of the variation in earnings management remains unexplained by the included variables. However, it is important to note that such low R-squared values are common in cross-sectional earnings management studies, as the complex and multifaceted nature of earnings management is difficult to fully capture in a single model. While the explanatory power of the model is limited, it is nonetheless consistent with the range of values typically observed in this research context. Despite the low R-squared, the model still provides valuable insights into the relationship between the selected variables and earnings management, albeit with the acknowledgment that there are likely other important factors at play that are not accounted for in this specification.

The substantial gap in explanatory power between the two channels reflects the mechanical relationship between the abnormal-CFO component of Total_REM and the operational controls, in contrast with the discretionary character of accrual manipulation that is less constrained by observable firm characteristics.

To examine whether the aggregate null masks heterogeneity across the three components of ESG, we re-estimate the baseline OLS specification with the Environmental, Social, and Governance pillar scores entered separately in place of the aggregate ESG variable. The pillar-disaggregated regressions, reported in Appendix B, reveal a more nuanced pattern. The Environmental pillar carries a marginally negative and statistically significant coefficient on AEM, while the Social and Governance pillars are statistically indistinguishable from zero in both the AEM and REM regressions. The Environmental-only AEM finding qualifies the aggregate null and provides partial support for H1.a along a specific channel of CSR engagement. This pillar-level pattern contrasts with the Italian evidence of Ricapito (2024), who locates the constraining effect of ESG on earnings management primarily in the Social and Governance pillars. The divergence may reflect the broader European geographic scope of our sample, in which the heightened materiality of environmental disclosure under the European regulatory environment makes the Environmental pillar a particularly salient channel through which CSR engagement disciplines accrual-based reporting choices relative to the Italian setting. The Environmental result is also coherent with the heightened materiality of environmental disclosure under the European regulatory environment, in which environmental compliance constitutes a particularly salient dimension of stakeholder pressure relative to the social and governance pillars.

5.2. The Joint-Substitution Specification

The baseline OLS regressions of Section 5.1 examine the AEM and REM channels separately. As Zang (2012) and Cohen and Zarowin (2010) emphasize, however, managers treat the two channels as strategic substitutes, and the estimation of an unconditional CSR coefficient on either proxy may be biased if firms that respond to CSR pressure by reducing AEM simultaneously increase REM, or vice versa. To address this concern, we re-estimate the H1 regressions on the strict overlapping sample of 1,098 firm-years described in Panel B of the Section 4 descriptive statistics, adding the alternative earnings management proxy as a cross-channel control on the right-hand side. The joint-substitution specification therefore regresses |DA_KM| on aggregate ESG, the Total_REM cross-control, the full control vector, and the saturated fixed effects, and symmetrically regresses Total_REM on ESG, the |DA_KM| cross-control, controls, and fixed effects. This is the same methodological device used by Kim et al. (2012) to disentangle accrual-based from real-activities earnings management in the U.S. setting.

The joint-substitution results yield two findings that materially refine the baseline picture. First, the cross-EM coefficients are positive in both regressions, with the |DA_KM| cross-control entering marginally positive in the Total_REM equation and the Total_REM cross-control entering positive but insignificant in the |DA_KM| equation. The positive sign of these cross-EM coefficients is informative. It indicates that, in the European cross-section, the two manipulation channels behave as complements rather than substitutes. Firms that engage in one form of earnings management tend to engage in the other as well. This pattern contrasts with the substitution dynamics that Zang (2012) and Cohen and Zarowin (2010) document in U.S. settings, where AEM and REM are typically deployed as alternative tools. The complementarity finding has direct implications for the interpretation of the H1 test, since it rules out the possibility that the aggregate null of Section 5.1 reflects an offsetting substitution between channels.

Second, once the cross-channel substitution control is included, the aggregate ESG coefficient on |DA_KM| becomes marginally negative and significant, with a point estimate of minus 0.003 and a p-value below the ten percent threshold (see Appendix D, Panel A). The corresponding ESG coefficient in the Total_REM equation remains statistically indistinguishable from zero. The emergence of a marginally negative ESG coefficient on AEM under the joint-substitution specification suggests that the baseline OLS regression of Section 5.1 understates the true CSR and AEM relationship by failing to absorb the variation in accrual choices that is mechanically correlated with simultaneous real-activities behaviour. The interpretation is that, conditional on the firm's level of real-activities manipulation, higher ESG performance is associated with a small but detectable reduction in accrual-based manipulation. This pattern is broadly consistent with the indivisibility of ethical conduct that Jones (1995) places at the heart of the moral track. Firms that build their business on trust and cooperation with stakeholders have an incentive to demonstrate that commitment consistently across all domains, including financial reporting, because stakeholders rely on the coherence of ethical behaviour across observable margins. Under this logic, the same managerial disposition that

produces substantive CSR engagement should also discipline the accrual choices through which reporting integrity is most directly observed by external parties. The magnitude of the AEM effect we recover is modest, which is consistent with the contextual contingency emphasised by Chih et al. (2008) and Liang and Renneboog (2017), but its directional sign and pillar-specific concentration support the Jones (1995) prediction that ethical commitment in the CSR domain extends to the financial-reporting domain rather than coexisting with opportunistic accruals. The pillar-disaggregated joint-substitution specification, also reported in Appendix B, strengthens the picture further: the Environmental pillar coefficient on AEM becomes more sharply negative under the substitution control, attaining significance at the five percent level. The Social and Governance pillars remain statistically indistinguishable from zero in both proxies.

To localise the channel-specific complementarity between AEM and REM to a single operational sub-channel, we re-estimate the Roychowdhury components separately. We regress each of the three abnormal-activity proxies, namely abnormal cash flow from operations (Abn_CFO), abnormal production costs (Abn_PROD), and abnormal discretionary expenses (Abn_DISEXP), on aggregate ESG, the |DA_KM| cross-control, the full control vector, and the saturated fixed effects. The results are reported in Appendix C. The aggregate ESG coefficient is statistically indistinguishable from zero in all three component regressions, mirroring the null on Total_REM. The |DA_KM| cross-control, however, loads positively and significantly on Abn_PROD, while remaining insignificant on Abn_CFO and Abn_DISEXP. This decomposition pinpoints the cross-channel complementarity to the production-cost dimension of real-activities manipulation. Firms that manage accruals also tend to manipulate production costs, but they do not systematically manage sales (Abn_CFO) or discretionary expenditures (Abn_DISEXP) in tandem. The pattern is coherent with the cost ranking that Roychowdhury (2006) and Graham et al. (2005) establish, in which overproduction is the lowest-cost form of real-activities manipulation and therefore the most natural complement to accrual-based discretion.

5.3. Causal Evidence from the NFRD Quasi-Natural Experiment (H2.a and H2.b)

The OLS and joint-substitution regressions of Sections 5.1 and 5.2 establish the conditional cross-sectional association between CSR and earnings management, but cannot speak to the causal direction emphasised by Christensen et al. (2021) in their review of mandatory CSR disclosure. Moving from association to causation requires an identification strategy that isolates an exogenous source of variation in firm-level ESG performance. We exploit the 2017 implementation of the EU Non-Financial Reporting Directive (NFRD) as a quasi-natural experiment that imposed binding mandatory ESG reporting obligations on European public-interest entities, following the heterogeneous-exposure logic developed by Fiechter et al. (2022). The compliance bite of the NFRD falls disproportionately on firms with the weakest pre-directive ESG performance, providing the cross-sectional differential exposure that supports a two-stage least-squares (2SLS) framework with firm and year fixed effects. The instrument is the interaction of the bottom-tercile pre-NFRD industry-adjusted ESG indicator (Treat) with a post-2017 dummy (Post), as defined in the Data section. Figure 1 of Section 4

presents the distribution of the pre-NFRD industry-adjusted ESG score across the 380 firms with usable pre-treatment data, with the bottom-tercile cutoff marked at minus 5.24.

5.3.1. First-Stage Evidence: The NFRD as a Binding Compliance Shock

The first-stage regression of contemporaneous aggregate ESG on the Treat times Post instrument, the firm-level control vector, and the firm and year fixed effects yields a coefficient of 11.43 on the instrument (cluster-robust standard error of 2.16, significant at the one percent level), as reported in column (1) of Table 9. Treated firms, defined as those in the bottom tercile of the pre-NFRD industry-adjusted ESG distribution, gained on average 11.43 ESG points more than control firms following the directive's 2017 effective date. The magnitude corresponds to approximately 0.6 standard deviations of the aggregate ESG distribution and provides quantitative confirmation of the visual evidence in Figure 2 of Section 4, which shows the steep post-2017 acceleration of the treated group's ESG trajectory. The first stage therefore confirms that the NFRD operated as a binding compliance constraint on the firms with the greatest pre-treatment scope to improve, in line with the heterogeneous-bite mechanism established by Fiechter et al. (2022) and the equity-market evidence of Grewal et al. (2019), who document that market reactions to the directive's passage concentrated in firms with weak pre-directive ESG performance. Among the first-stage controls, analyst coverage carries a positive and significant coefficient, consistent with the monitoring channel emphasised by He et al. (2022) and Sun et al. (2024), through which heightened external scrutiny amplifies firms' responsiveness to mandatory disclosure obligations.

Table 9. Causal Effect of ESG on Earnings Management: Two-Stage Least Squares

Variable	**Stage 1: ESG**	**Stage 2: AEM**	**Stage 2: REM**
Treat × Post (IV)	11.427***		
	(2.156)		
ESG Score (instrumented)		-0.002	-0.007
		(0.005)	(0.005)
Size (log Mkt Cap)	-0.872	0.005	0.053*
	(0.560)	(0.094)	(0.024)
Market-to-Book	0.078	0.015	-0.006
	(0.160)	(0.011)	(0.004)
ROA	-3.317	-0.032	0.153
	(3.965)	(0.264)	(0.100)
Leverage	-2.928	-0.076	0.299
	(3.948)	(0.237)	(0.216)

CFO (scaled)	-7.050	0.098	-0.549**
	(5.119)	(0.740)	(0.207)
log(1+Firm Age)	5.857	-0.216	-0.184
	(3.929)	(0.236)	(0.134)
Analyst Coverage	0.377***	-0.006	-0.000
	(0.091)	(0.006)	(0.003)
Big4 Auditor	-1.180	0.074	0.064
	(1.013)	(0.045)	(0.037)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	1,811	986	847
R-squared	0.808	0.368	0.779
First-stage F	134.135	134.135	103.543
Kleibergen-Paap rk Wald F	4.315	4.315	3.600
Wu-Hausman p-value	-	0.9025	0.0080

Notes: Column (1) reports the first-stage regression of ESG on the NFRD instrument Treat × Post (Treat = bottom tercile of pre-NFRD industry-adjusted ESG; Post = 1 if year >= 2017). Columns (2)-(3) report the second-stage 2SLS estimates with AEM (|DA_KM|) and REM (Total_REM) as outcomes. All specifications include firm and year fixed effects with two-way clustered standard errors. Stock and Yogo (2005) weak-instrument critical values: F > 10 (conventional); F > 23.11 (5% maximum-bias). The first-stage F-statistic is computed under the cluster-robust standard-error structure (firm and year) used throughout the analysis. The Kleibergen-Paap rk Wald F is computed under heteroskedasticity-only standard errors and is reported as a stricter benchmark. The divergence between the two statistics reflects the difference in inference assumptions, not weak identification under the preferred cluster-robust specification. The Wu-Hausman p-value tests the null that ESG is exogenous in the structural equation. Rejection (low p-value) indicates 2SLS is required.

5.3.2. Second-Stage Evidence: The Null Causal Effect on Both Manipulation Channels

The second-stage 2SLS regressions take the predicted exogenous component of ESG as the main regressor and the two earnings management proxies as the dependent variables in turn. The structural coefficients on instrumented ESG are reported in columns (2) and (3) of Table 9. The point estimate is minus 0.002 in the AEM regression (standard error 0.005, N equal to

986) and minus 0.007 in the REM regression (standard error 0.005, N equal to 847). Both estimates are statistically indistinguishable from zero, and the upper bounds of their 95 percent confidence intervals correspond to economically negligible effect sizes. The structural evidence therefore provides no causal support for the disciplining prediction of H2.a, under which the NFRD-induced increase in CSR engagement should reduce the magnitude of opportunistic earnings management through the joint operation of the ethical-internalisation channel emphasised by Christensen et al. (2021) and the external-monitoring channel emphasised by He et al. (2022), Dhaliwal et al. (2011), and Xue et al. (2025).

The convergence of the second-stage causal estimate and the OLS baseline estimate of Section 5.1 deserves emphasis. Both methodologies produce coefficients close to zero on the AEM regression. The instrumental variable estimate of minus 0.002 lies within standard-error distance of the OLS baseline, while the REM IV estimate of minus 0.007, although nominally larger in absolute value, remains statistically indistinguishable from zero. This convergence between observational and quasi-experimental estimates strengthens the inference that the absence of a CSR and earnings management relationship in the European panel reflects a genuine economic phenomenon rather than the failure of one estimation approach to address endogeneity. The Wu-Hausman test reported in Table 9 supports a more granular reading. For the AEM channel, the test does not reject the null that ESG is exogenous in the structural equation ($p = 0.90$), which is consistent with the close numerical agreement between the OLS and 2SLS estimates and indicates that the OLS baseline is not materially biased by endogeneity in this channel. For the REM channel under firm and year fixed effects, by contrast, the test rejects exogeneity at the one percent level ($p = 0.008$), providing direct statistical support for the use of 2SLS as the appropriate identification strategy for real-activities manipulation. The REM IV estimate therefore represents the corrected causal coefficient rather than a complementary check, although its substantive conclusion is unchanged. The point estimate is small, statistically indistinguishable from zero, and falls well within the conventional null region. We retain the complementary framing advocated by Larcker and Rusticus (2010) for the AEM channel and treat the REM 2SLS as the preferred estimator on the strength of the Wu-Hausman evidence. The marginally significant Environmental-pillar OLS finding documented in Section 5.1 therefore appears to capture a cross-sectional association rather than a causal mechanism transmitted through the NFRD compliance shock. Among the second-stage controls, the Size coefficient in the REM regression is marginally positive, and the CFO coefficient is significantly negative, while the remaining controls do not attain conventional significance.

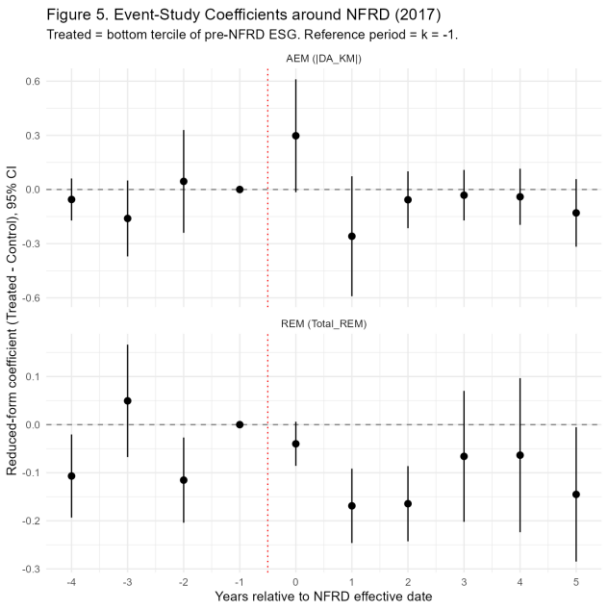
The reduced-form difference-in-differences regression, which directly regresses each earnings management outcome on the Treat times Post interaction and the saturated controls, corroborates the second-stage null. The reduced-form coefficient is small and statistically insignificant in both the AEM and REM regressions (see Appendix D, Panel B), and the implied Wald-ratio IV coefficients match the joint 2SLS estimates from Table 9 up to small differences arising from the precise sample composition. The reduced-form null is informative in its own right. It indicates that the NFRD did not produce a detectable level shift in either earnings management proxy among the firms most exposed to its compliance bite, even before

any structural transformation through instrumented ESG. Figures 3 and 4 of Section 4 illustrate this graphically by plotting the mean of |DA_KM| and Total_REM by treatment group over time, showing that no sustained divergence opens up between the two groups around the 2017 effective date. From the perspective of H2.a, the directive successfully altered what firms report along the CSR dimension, as evidenced by the strong first stage, but it did not translate this disclosure shock into measurable improvements in financial reporting quality along either the accrual-based or the real-activities channel.

The structural null is in tension with Fiechter et al. (2022), who document substantial NFRD-induced changes in firms' operational behaviour. Two reconciling factors are worth noting and are grounded in the literature reviewed in Section 2. First, the outcomes that Fiechter et al. (2022) examine are measured at a different level of granularity than the continuous earnings management proxies used here, and may therefore capture margins of managerial response that the |DA_KM| and Total_REM proxies do not. Second, the sample window of this study extends through 2025, well beyond the post-NFRD adjustment period emphasised in early evaluations of the directive. If the directive's effects on financial reporting behaviour were concentrated in the immediate transition window and dissipated as compliance practices matured, then long-window estimates would naturally attenuate toward zero. The observation by Xue et al. (2025) that the mere existence of a mandatory CSR report ceases to operate as a credible signal once compliance becomes universal is consistent with the long-run attenuation interpretation: the signalling function highlighted by Dhaliwal et al. (2011) is neutralised under universal mandatory disclosure, and the disciplining residual proves insufficient to generate a measurable reporting-quality effect in the aggregate.

5.3.3. Event-Study Diagnostics

The 2SLS framework rests on the assumption that treated and control firms would have followed parallel trends in earnings management absent the NFRD. We assess this assumption with a dynamic event-study specification that interacts the Exposure indicator with event-time dummies, with the immediate pre-treatment year omitted as the reference period. Figure 5 below plots the dynamic event-study coefficients for both the AEM and REM outcomes. The event-study coefficients for the AEM outcome are statistically indistinguishable from zero in the pre-treatment leads, supporting the parallel-trends assumption for the accrual-based channel. The post-treatment path returns to zero after a borderline positive spike at the treatment date, a pattern consistent with a one-time disclosure-related disruption rather than a sustained change in accrual behaviour. The event-study coefficients for the REM outcome, by contrast, show statistically significant negative pre-trends at event times minus four and minus two, indicating that treated and control firms were already on diverging real-activities trajectories before the directive took effect. The pre-trend violation in the REM panel weakens any causal interpretation of the REM-specific 2SLS estimate. The AEM coefficient retains its identifying assumption, and the REM coefficient should be interpreted with caution. The event-study evidence is consistent with the finding that the REM cross-sectional correlation observed in Section 4 weakens substantially in the strict overlapping sample, suggesting that part of the apparent REM association is mechanical rather than behavioural.



A placebo test imposing a counterfactual treatment date in 2013, within the pre-NFRD subsample, produces a statistically insignificant first-stage coefficient of 2.32 with a standard error of 1.77 (see Appendix D, Panel C). The placebo first-stage is roughly a fifth of the magnitude of the genuine 2017 first-stage of 11.43 and lies well within standard-error distance of zero. The falsification evidence confirms that no exogenous CSR shock is detectable at the placebo date, lending additional credibility to the identification strategy.

5.3.4. Evidence on the Substitution Prediction (H2.b)

Hypothesis H2.b, building on the post-SOX evidence of Cohen et al. (2008) and the contemporary European evidence of Fiechter et al. (2022), predicts that mandatory CSR disclosure should generate a substitution from AEM to REM as managers pivot toward less detectable manipulation channels under heightened scrutiny. The empirical signature of H2.b is a negative reduced-form coefficient on AEM coupled with a positive reduced-form coefficient on REM. The reduced-form regressions in Section 5.3.2 do not display this signature. The AEM coefficient is essentially zero, and the REM coefficient is small and equally insignificant. Neither sign nor significance pattern matches the H2.b prediction. Combined with the cross-sectional complementarity finding of Section 5.2, the data are inconsistent with the strong substitution prediction of H2.b in the European setting.

The heterogeneous treatment effect analysis, reported in Table 27, reveals that the NFRD's impact on earnings management is strongly concentrated in the firms with the very lowest pre-directive CSR performance. Estimating the 2SLS specification across increasingly narrow exposure definitions shows a statistically significant negative effect on AEM for firms in the bottom 20% and 25% of the industry-adjusted pre-NFRD ESG distribution, while no significant effect is detected for the broader bottom-tercile or bottom-half samples. This pattern suggests that the directive's disciplining effect on accrual-based earnings management is largely driven by the subset of firms facing the most acute compliance pressure relative to their industry peers. The implication is that the NFRD's impact on reporting behavior is not uniform,

but rather concentrated in the left tail of the ex ante CSR distribution where the regulatory burden of mandatory disclosure is most binding.

One reconciling possibility, grounded in the temporal dynamics emphasised by Cohen et al. (2008) for the post-SOX transition, is that the AEM-to-REM substitution operates principally during the immediate post-mandate adjustment window and dissipates as firms adapt to the new disclosure regime. The 2010 to 2025 panel used in this study averages the potential immediate substitution effect with longer-run reversion, which may yield null aggregate estimates even if a transitional substitution occurred. A second reconciling factor is that the European institutional environment, characterised by the universal applicability of NFRD obligations across in-scope firms, may neutralise the asymmetric monitoring intensity that Cohen et al. (2008) identified as the driver of post-SOX substitution. When mandatory CSR reporting becomes universal, both AEM and REM are exposed to the same uplift in external scrutiny, and the relative-detection-cost asymmetry that motivates the substitution mechanism is attenuated. This logic aligns with the observation by Xue et al. (2025) that universal mandatory reporting neutralises the signalling function of CSR disclosure.

5.4. Robustness Analyses

The 2SLS results are subjected to **six** robustness specifications that vary the treatment definition, the sample composition, and the control vector. Each robustness regression re-estimates the second-stage structural coefficient on instrumented ESG for both the AEM and REM outcomes, holding the firm and year fixed effects and the full control vector constant. The six specifications are reported as separate rows in Table 10. Two additional robustness analyses, the event-study diagnostics and the placebo test specified in Section 3.3.7, are reported separately: the event-study coefficients in Section 5.3.3 (Figure 5) and the placebo first-stage in Appendix D, Panel C.

Table 10. Robustness of the Causal Estimate

Specification	**Coef (AEM)**	**SE (AEM)**	**N (AEM)**	**F (AEM)**	**Coef (REM)**	**SE (REM)**	**N (REM)**	**F (REM)**
Main 2SLS (bottom tercile, firm + year FE)	-0.002	(0.005)	986		-0.007	(0.005)	847	
Bottom quartile cutoff	0.001	(0.011)	986		-0.001	(0.005)	847	
Bottom-half cutoff	0.001	(0.006)	986		-0.006	(0.004)	847	
Continuous intensity instrument	0.005	(0.007)	986		-0.005	(0.003)	847	

IPTW re-weighting on observables	-0.005	(0.004)	839	-0.003	(0.005)	750
Excluding Big4 from controls	-0.007	(0.009)	1,222	-0.003	(0.004)	1,031
Restricted to mixed-treatment industries	-0.001	(0.005)	900	-0.007	(0.006)	758

*Notes: Each row reports the second-stage coefficient on instrumented ESG from a separate 2SLS specification. Coef columns report the estimate with significance stars; SE columns the cluster-robust standard error; N the estimation sample size; F the first-stage F-statistic on the excluded instrument. The first row reproduces the main estimate from Table 10. Stars: *** p<0.01, ** p<0.05, * p<0.10.*

The first three robustness specifications vary the definition of the treatment group. The first specification replaces the bottom-tercile cutoff with the bottom quartile of the pre-NFRD industry-adjusted ESG distribution, narrowing the treated population to the most weakly performing firms. The second-stage AEM coefficient is essentially zero (point estimate of 0.001, standard error of 0.011), and the REM coefficient is similarly null (minus 0.001, standard error of 0.005). The second specification widens the treatment threshold to the bottom-half cutoff, doubling the treated population. The AEM coefficient remains close to zero (0.001, standard error of 0.006) and the REM coefficient is moderately negative but statistically insignificant (minus 0.006, standard error of 0.004). The third specification replaces the binary treatment indicator with a continuous-intensity instrument equal to one minus the firm's pre-treatment industry-adjusted ESG percentile rank, exploiting the full distribution of pre-treatment exposure rather than a discrete tercile cutoff. The AEM coefficient becomes marginally positive at 0.005 (standard error of 0.007), and the REM coefficient is moderately negative at minus 0.005 (standard error of 0.003). None of these three specifications produces a statistically significant second-stage estimate, and the point estimates remain within a narrow band of the main 2SLS estimates of minus 0.002 and minus 0.007 for the AEM and REM channels respectively. The stability across treatment-definition variants indicates that the headline null is not an artefact of the particular tercile cutoff used in the main analysis. It also confirms that the identification does not depend on the discrete-versus-continuous parameterisation of pre-treatment exposure, which strengthens the inference that the underlying causal effect is genuinely small in the European setting.

The fourth robustness specification applies inverse-probability-of-treatment weights to address the pre-treatment imbalance documented in Table 6 of Section 4 on size, market-to-book, analyst coverage, log_Age, and Big4 audit. The propensity-score reweighting re-balances treated and control firms on these pre-treatment observables before re-estimating the 2SLS specification on the weighted sample of 839 (AEM) and 750 (REM) firm-years. The IPTW-

weighted AEM coefficient is minus 0.005 (standard error of 0.004), and the IPTW-weighted REM coefficient is minus 0.003 (standard error of 0.005). Both estimates remain statistically insignificant and economically close to the unweighted estimates from Table 10. The marginal increase in the absolute value of the AEM coefficient under IPTW reweighting is not statistically meaningful but suggests that, in the rebalanced sample, the AEM channel shows a directional sign consistent with H1.a, again without attaining significance. The substantive implication is that the pre-treatment imbalance on observables documented in Section 4 is not driving the main null. The headline finding survives the methodological correction for selection on observables, providing reassurance that the structural estimate is not contaminated by the size, monitoring, and audit-quality imbalances visible in the pre-treatment balance test.

The fifth robustness specification excludes the Big4 indicator from the control vector, recovering the larger auditor-data sample of 1,222 firm-years for AEM and 1,031 firm-years for REM. This specification addresses the concern that the relatively high missingness on the Big4 variable in the main sample may be selecting on a non-random subset of firm-years. The AEM coefficient under this specification is minus 0.007 (standard error of 0.009), and the REM coefficient is minus 0.003 (standard error of 0.004). Neither coefficient is statistically significant, and the increase in the AEM point estimate is offset by a roughly proportional increase in its standard error, so the t-statistic remains comparable. The substantive conclusion is that the missingness in the Big4 variable does not threaten the validity of the headline inference. The directional sign of the AEM coefficient continues to be negative in this expanded sample, consistent with the directional pattern that emerged under the joint-substitution specification of Section 5.2.

The sixth robustness specification restricts the estimation sample to NAICS subsectors that contain both treated and control firms, removing the subsectors that are entirely composed of one or the other treatment group. The motivation for this restriction is methodological. In the full sample, identification of the 2SLS coefficient could, in principle, exploit cross-industry variation in treatment intensity, which may confound the within-industry causal effect with industry-level heterogeneity in earnings management behaviour. Restricting to mixed-treatment subsectors removes this concern by forcing identification to operate within industries that contain both groups. The restricted-sample AEM coefficient is essentially zero (minus 0.001, standard error of 0.005), and the REM coefficient is minus 0.007 (standard error of 0.006). Both estimates remain statistically insignificant and substantively indistinguishable from the main estimates. The within-industry identification therefore produces the same null finding as the full-sample identification, ruling out the interpretation that the headline result is driven by cross-industry confounding.

Taken together, the six robustness specifications reported in Table 10 deliver a consistent message. The second-stage coefficient on instrumented ESG remains confined to a narrow band running from minus 0.007 to plus 0.005 across the AEM regressions, and from minus 0.007 to minus 0.001 across the REM regressions. In no specification does either coefficient approach the conventional ten percent significance threshold. The robustness battery rules out the specific tercile cutoff, the discrete treatment parameterisation, the pre-treatment imbalance

on observables, the missingness on auditor data, and cross-industry confounding as alternative explanations for the headline null. The stability of the structural estimate across these six variants, together with the event-study and placebo diagnostics reported in Section 5.3.3 and Appendix D respectively, substantially strengthens the inference that the null causal effect of NFRD-induced CSR engagement on European earnings management reflects a genuine economic phenomenon rather than a methodological artefact specific to the main specification.

5.5. Synthesis of Findings and Discussion

The empirical evidence assembled in Sections 5.1 through 5.4 admits a coherent interpretation against the four hypotheses developed in Section 2.3 of the literature review. The Transparent Financial Reporting Hypothesis (H1.a), which predicts that higher ESG performance is associated with a lower magnitude of earnings management across both the accrual-based and real-activities channels, receives only weak and partial support in the European panel. The baseline OLS regressions of Section 5.1 produce statistically null coefficients on aggregate ESG in both the AEM and REM equations. The joint-substitution specifications of Section 5.2, which hold the alternative manipulation channel constant, recover a marginally negative ESG coefficient on AEM and a more sharply negative Environmental-pillar coefficient on AEM, but the REM channel remains uniformly null across all specifications. The headline verdict on H1.a is therefore that the directional sign of the OLS evidence is consistent with the moral track logic of Jones (1995) and the U.S. evidence of Kim et al. (2012), but the magnitudes are small, the effect is channel-specific (limited to AEM), pillar-specific (concentrated in the Environmental dimension), and only detectable under the joint-substitution specification. The Opportunistic Financial Reporting Hypothesis (H1.b), which predicts a positive association between ESG and earnings management under the moral-licensing logic of Ormiston and Wong (2013) and Wadhaad et al. (2023), is rejected: no specification produces a positive coefficient on ESG in either channel.

The Ethical Internalisation and External Monitoring Hypothesis (H2.a), which predicts that the NFRD-induced increase in mandatory ESG engagement should reduce the overall magnitude of earnings management through the joint operation of the internalisation channel of Christensen et al. (2021) and the external monitoring channel of He et al. (2022) and Xue et al. (2025), is rejected by the 2SLS evidence of Section 5.3. The directive successfully induced a substantial differential improvement in ESG performance among treated firms, with a first-stage coefficient of 11.43 ESG points, but this exogenously induced CSR variation does not translate into any detectable reduction in either accrual-based or real-activities manipulation. The marginally negative AEM finding documented in the OLS substitution model therefore appears to capture a cross-sectional association rather than a causal mechanism transmitted through the NFRD compliance shock. The reduced-form difference-in-differences regressions corroborate the second-stage null, with insignificant Treat times Post coefficients on both proxies. The event-study diagnostics of Section 5.3.3 support the identifying assumption for the AEM channel and weaken it for the REM channel, which is a further qualification on the strength of the REM inference but does not change the substantive AEM conclusion.

The Strategic Substitution Hypothesis (H2.b), which predicts an AEM-to-REM pivot under the heightened scrutiny induced by mandatory CSR reporting, is also rejected. The reduced-form coefficients on AEM and REM are both small, both statistically indistinguishable from zero, and crucially do not display the directional pattern of a decrease in AEM coupled with an increase in REM that the substitution mechanism of Cohen et al. (2008) would imply. The cross-sectional complementarity finding of Section 5.2, in which AEM and REM behave as complements rather than substitutes in the European baseline, runs directly counter to the H2.b prediction even outside the regulatory-shock window. The robustness specifications of Section 5.4 confirm the stability of the null causal estimates across six alternative parameterisations.

Several possible reasons for the predominantly null findings deserve discussion, all of which are grounded in the literature reviewed in Section 2. The first reason is institutional. As Chih et al. (2008) and Liang and Renneboog (2017) emphasize, the CSR and earnings management relationship is highly contingent on the institutional environment in which firms operate. The civil-law jurisdictions that compose the European Union foster a higher baseline of CSR engagement as a matter of institutional compliance rather than as a discretionary ethical signal. In such an environment, observed variation in ESG scores may reflect compliance gradations rather than meaningful differences in managerial integrity, which would weaken the moral-track mechanism that Kim et al. (2012) documented in the U.S. setting, where voluntary CSR engagement carries more informational content. A second reason is the universality of the NFRD mandate. Following the logic of Xue et al. (2025), when mandatory ESG reporting becomes a universal legal requirement, the signalling function of CSR disclosure is neutralised. Stakeholders can no longer use the mere existence of a CSR report to identify an ethical manager. The disciplining residual of the mandate proves insufficient to generate detectable reporting-quality effects once the signalling channel is removed. A third reason is the non-monotonic structure of the ESG and earnings management relationship documented by Elrazaz and Aljifri (2026), who find that the constraining effect of ESG on opportunism operates only within specific performance thresholds rather than across the full distribution. Averaging across the distribution in our linear specifications may mask offsetting effects at the tails, contributing to the aggregate null. A fourth reason relates to measurement. The construction of the |DA_KM| and Total_REM proxies inherit the well-known limitations of the underlying Kothari et al. (2005) and Roychowdhury (2006) models. As Dechow et al. (2012) and Dechow and Skinner (2000) caution, traditional accruals models struggle to separate genuine business growth from opportunistic discretion, and the residuals of cross-sectional models contain a non-trivial component of estimation noise that attenuates true coefficients toward zero. The pre-trend violation in the REM event-study panel, and the disappearance of the bivariate ESG and REM correlation in the strict overlapping sample documented in Section 4, both suggest that part of the REM signal is mechanical rather than behavioural. The combination of these institutional, regulatory, methodological, and measurement factors provides a coherent explanation for the predominantly null findings reported in this section, and motivates the broader implications and limitations discussed in the concluding section.

6. Conclusion

This thesis has examined whether Corporate Social Responsibility constrains earnings management among European listed firms, drawing together a cross-sectional OLS analysis of the conditional association between ESG performance and the two principal channels of earnings management with a two-stage least-squares design that exploits the 2017 NFRD as a quasi-exogenous shock to identify the causal effect of mandatory CSR engagement on financial reporting quality. Read across the four hypotheses developed in Section 2, the European evidence converges on a coherent overall message. The directional prediction of the Transparent Financial Reporting Hypothesis is recovered only along a narrow margin, concentrated in the Environmental pillar of ESG and in the accrual-based channel of manipulation under a joint-substitution specification, while the Opportunistic Financial Reporting Hypothesis finds no support anywhere in the cross-section. The NFRD, although effective in raising the ESG performance of ex ante low-CSR firms by a substantial margin, does not translate this regulatory shock into measurable improvements in financial reporting quality through either the disciplining channel anticipated by H2.a or the channel-substitution mechanism anticipated by H2.b. The two manipulation channels behave, in the European setting, as complements rather than as strategic substitutes as reported in some studies for the firms from the United States.

The main contribution of these findings is not to confirm the "moral track" argument from earlier work, but to clarify when and where that argument is likely to hold, and to introduce two angles that, to our knowledge, have not been examined jointly in the existing literature.

First, this study is the first to exploit the NFRD as a regulatory shock to test the CSR-earnings management relationship in a causal framework. Prior research has overwhelmingly relied on observational, cross-sectional, or panel designs in which CSR engagement is treated as endogenous to firm characteristics, and identification rests on controls rather than on exogenous variation. By exploiting the 2017 effective date of the NFRD as a sharp regulatory break with heterogeneous compliance bite across firms with different pre-existing CSR levels, we move the literature beyond correlational evidence and provide one of the first causal estimates of how a CSR-disclosure mandate affects financial reporting behaviour.

Second, we explicitly trace the chain from mandatory disclosure to CSR performance to financial reporting quality, a linkage that the prior literature has discussed conceptually but not tested as an integrated mechanism. Existing work tends to examine either the disclosure-performance link (does mandatory reporting change what firms actually do on CSR?) or the performance-earnings management link (do socially responsible firms manage earnings less?), but rarely both within the same identification strategy. By estimating these stages together, we show where the chain breaks: NFRD changed what firms report and, to some extent, how they behave on CSR, but this did not propagate into the financial reporting margin in the way the moral-track hypothesis predicts. Identifying where the transmission fails is itself informative, because it tells regulators and researchers that disclosure mandates and financial reporting integrity may operate through more independent channels than the theory assumes.

Third, the European null results help discipline the broader literature. In Europe, three factors plausibly explain the absence of an effect: (1) civil-law institutions compress the cross-firm variation in reporting discretion that the moral-track mechanism needs to operate on; (2) universal mandatory disclosure weakens CSR's signalling value, making it harder to distinguish genuine engagement from cosmetic compliance; and (3) the underlying relationship is likely non-monotonic, so linear specifications can mask offsetting effects. Together, these points suggest that the largely null results in post-NFRD Europe are not a refutation of the international evidence but a boundary condition on it. The CSR-earnings management link appears to be conditional on an institutional setting in which CSR carries informational content, which mandatory regimes partly eliminate.

For regulators, the implications are twofold. The NFRD demonstrably changed what firms report on CSR, but in the post-2017 period, it did not change how firms report financially in the direction the moral-track view would predict. Policymakers should therefore be cautious about expecting spillover benefits from sustainability-disclosure regimes onto financial reporting quality. The two appear to be governed by largely separate incentives and enforcement structures. This has direct relevance for the ongoing CSRD rollout, which deepens and broadens the NFRD's requirements, which is that stronger sustainability reporting should not be assumed to deliver, as a by-product, cleaner earnings.

Taken together, the study's contribution is substantive, methodological, and policy-relevant. Substantively, it shows that the conditional patterns documented in international observational work do not extend mechanically to a mandatory-disclosure European setting. Methodologically, it brings causal identification to a literature that has been largely correlational and integrates two stages of the CSR-reporting chain that have previously been studied in isolation. For policy, it cautions against treating sustainability-disclosure regulation as an indirect lever on financial reporting quality. Established findings from observational studies are left intact, but their scope is now more clearly delineated.

6.1 Limitations

Several limitations qualify the inferences drawn in this study and warrant explicit acknowledgement. The first concerns the identification strategy. The absence of a clean external control group, such as comparable non-EU firms unaffected by the NFRD, means that identification rests entirely on within-EU heterogeneity in treatment intensity. The structural coefficient therefore captures the differential impact of more aggressive compliance among treated firms rather than the absolute effect of the regulation, and the LATE interpretation identifies the effect on compliers rather than the average effect across all firms. The just-identified nature of the 2SLS specification precludes overidentification tests, so the case for exclusion rests on institutional argumentation rather than statistical testing. The event-study diagnostics further reveal a pre-trend violation in the real-activities panel that weakens the REM-specific causal interpretation, while the accrual-based channel retains its identifying assumption.

Closely related is a more fundamental limitation of the IV approach itself: of the three conditions required for instrument validity, only relevance is directly testable from the data. The strong first-stage coefficient and the associated F-statistic provide statistical evidence on this condition, but exclusion and exogeneity must be defended on institutional grounds rather than verified empirically. The principal threat to the exclusion restriction is that the NFRD plausibly improved firms' broader information environment, through internal controls, mandatory risk-management disclosures, and intensified external scrutiny, in ways that could constrain manipulation independently of substantive CSR change. The analyst-coverage and Big-Four controls partially absorb the monitoring channel, and the firm fixed effects absorb time-invariant differences in baseline disclosure infrastructure, but these mitigations do not eliminate the concern. The Wu-Hausman test reported in Table 9 rejects exogeneity for the REM channel at conventional significance levels, providing direct statistical support for the IV strategy in that channel. For the AEM channel, by contrast, the test fails to reject, and we therefore treat the AEM IV results as complementary to the OLS baseline following Larcker and Rusticus (2010), since failure to reject can reflect insufficient power as well as genuine exogeneity.

The second concerns measurement. The earnings management proxies inherit the well-documented difficulties of the underlying Kothari et al. (2005) and Roychowdhury (2006) models in separating legitimate growth-driven accruals and operational decisions from opportunistic discretion, and the residuals of cross-sectional models contain estimation noise that attenuates true coefficients toward zero. A related concern is that the cross-sectional estimation of the EM models at the NAICS subsector-year level requires a minimum of ten observations per cell, and thinly populated industry cells generate noisier non-discretionary benchmarks against which discretionary residuals are measured. The sensitivity analysis at stricter thresholds of 15 and 20 observations per cell reported in Appendix F confirms that, while the |DA_KM| proxy contains a noise component that attenuates as the threshold tightens, the headline H1 ESG coefficient is unaffected because the noise component is additive to the residual rather than correlated with the regressor of interest. The CSR construct itself, proxied by the aggregate LSEG Workspace ESG score, conflates substantive corporate practices with disclosure quality, a limitation that the pillar-disaggregated analysis partly mitigates but does not eliminate. The high missingness rate on the auditor identification field, unpopulated for approximately three-quarters of cleaned firm-years, is addressed in robustness checks dropping the Big-Four control, but the underlying data limitation remains.

The third concerns sample composition and temporal scope. The pre-treatment imbalance documented in Section 4 on size, monitoring, market-to-book ratio, scaled cash flow, log firm age, and Big-Four audit incidence is addressed through the inclusion of these variables as controls and through inverse-probability-of-treatment reweighting, but selection on unobserved firm characteristics correlated with both treatment and earnings management cannot be ruled out. A further constraint is the size of the analytical sample itself. While the baseline regression panel retains 1,944 distinct firms after applying the listing, data-availability, and industry-cell filters described in Section 3, the NFRD analytical sample on which the causal estimates are identified is considerably smaller: 223 firms (2,836 firm-years)

survive the additional pre-treatment-anchor and balanced-panel requirements, and the second-stage 2SLS regressions are estimated on subsamples of 986 firm-years (AEM) and 847 firm-years (REM) after listwise deletion on the full control vector. This is only modestly above the minimum that conventional power calculations would prescribe for detecting effect sizes of the magnitude reported in the international CSR-earnings management literature. Even if the second-stage coefficients had reached conventional levels of statistical significance, the resulting estimates would have rested on a sample whose statistical power leaves a limited margin for confident inference, and the null results reported here should accordingly be read as consistent with, rather than as decisive evidence against, a small underlying effect that a better-powered design could detect. The 2017 to 2025 post-treatment window, while long by the standards of regulatory-shock studies, averages the immediate adjustment effects emphasised in the post-Sarbanes-Oxley literature with longer-run reversion, and may not fully capture treatment dynamics that the successor Corporate Sustainability Reporting Directive (CSRD) is reshaping from 2024 onward.

6.2 Avenues for Further Research

Several avenues for further research follow naturally from the limitations enumerated above. Extension of the analysis to the CSRD period will permit a direct comparison of regulatory regimes and a test of whether the more granular and assurance-backed disclosure requirements of the successor directive alter the empirical signature documented here. The application of non-linear and threshold-based econometric specifications may reveal heterogeneity that the linear models employed in the present analysis necessarily mask, particularly at the tails of the ESG distribution where the moral-track mechanism is most likely to operate. Alternative measures of CSR substance, including third-party assurance indicators, materiality-weighted disclosure quality scores, and text-based measures derived directly from non-financial reports, would relax the dependence on commercial ESG ratings and permit a sharper separation of disclosure quality from underlying corporate practice. Comparative analyses pairing European firms with non-EU controls, where data availability permits, would supply the external counterfactual that the present design lacks and refine the LATE estimate into something closer to an average treatment effect. Finally, a deeper investigation of the pillar-specific Environmental finding, the only directional signal to emerge with any consistency from the European panel, may locate the precise mechanisms through which environmental disclosure interacts with accrual discretion and clarify whether the broader moral-track logic can be recovered under more demanding disclosure regimes or only through finer measurement of CSR substance itself.

Appendices

Appendix A. Regression equations used in Section 3.2

Eq.	Step	Specification
(1)	Step 1	$\|DA\ KM\|_{i,t}\ or\ Total\ REM_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \gamma' X^R{}_{i,t} + FE + \varepsilon_{i,t}$
(2)	Step 2	$\|DA\ KM\|_{i,t}\ or\ Total\ REM_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(3)	Step 3a	$\|DA\ KM\|_{i,t}\ or\ Total\ REM_{i,t} = \beta_0 + \beta_1 PILLAR_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(4)	Step 3b	$REM\ COMP_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(5)	Step 4	$\|DA\ KM\|_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \beta_2 Total\ REM_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(6)	Step 4	$Total\ REM_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \beta_2 \|DA\ KM\|_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(7)	Step 4	$\|DA\ KM\|_{i,t} = \beta_0 + \beta_1 PILLAR_{i,t} + \beta_2 Total\ REM_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(8)	Step 4	$Total\ REM_{i,t} = \beta_0 + \beta_1 PILLAR_{i,t} + \beta_2 \|DA\ KM\|_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(9)	Step 4	$REM\ COMP_{i,t} = \beta_0 + \beta_1 ESG_{i,t} + \beta_2 \|DA\ KM\|_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$
(10)	Step 4	$REM\ COMP_{i,t} = \beta_0 + \beta_1 Env_{i,t} + \beta_2 \|DA\ KM\|_{i,t} + \gamma' X_{\ i,t} + FE + \varepsilon_{i,t}$

Notes. The table presents the ten unique regression equations estimated in the four-step OLS framework of Section 3.2; substituting alternative dependent variables and CSR proxies into these equations yields the full set of estimated specifications.

Variable definitions. |DA_KM| is the absolute value of performance-adjusted Kothari discretionary accruals (the AEM proxy). Total_REM is the aggregate Roychowdhury real-earnings-management proxy, defined as Abn_PROD - Abn_CFO - Abn_DISEXP. REM_COMP denotes one of the three individual real-EM channels {Abn_CFO, Abn_PROD, Abn_DISEXP}, which appear separately as dependent variables in the channel-decomposition specifications (Eqs. (4), (9), and (10)). $ESG_{i,t}$ is the aggregate LSEG ESG score; PILLAR denotes one of the three constituent pillars {Environmental, Social, Governance}, and Env is the Environmental pillar alone, isolated in the final specification because the channel-decomposition results motivate a focused look at the environmental dimension.

Control vector. X denotes the full firm-level control vector {Size, MB, ROA, Lev, CFO_scaled, Log_Age, Analyst_Cov, Big4}; X^R denotes the reduced control vector used in the Step 1 large-sample sensitivity analysis, which excludes MB and Lev (the two controls with the highest data attrition).

Estimation. All specifications are estimated by ordinary least squares and include year and NAICS-subsector fixed effects (denoted FE in the equations). Standard errors are two-way clustered at the firm and year levels following Petersen (2009) and Gow et al.(2010). All continuous variables are winsorized at the 1st and 99th percentiles.

Step structure. Step 1 (Eq. 1) is a large-sample sensitivity check on the reduced control vector. Step 2 (Eq. 2) is the primary aggregate H1 test. Step 3a (Eq. 3) disaggregates the CSR proxy

into its three pillars; Step 3b (Eq. 4) decomposes the aggregate real-EM proxy into its three operational channels. Step 4 (Eqs. 5-10) implements the joint estimation discussed in Section 3.2.2, in which the equations are estimated on the strict overlapping sample with the alternative manipulation proxy included as a cross-control to address the strategic substitutability between accrual-based and real EM (Zang, 2012; Cohen, Dey, & Lys, 2008). The sign on the cross-control coefficient (β_2) is itself informative: a positive value indicates complementarity (joint use of both channels), a negative value indicates substitution (managers trading one channel against the other).

Appendix B. Earnings Management by CSR Pillars

Variable	AEM	REM
Panel A. Aggregate ESG (Main Table 8 reproduced for comparison)		
Dependent variable: |DA_KM| / Total REM		
ESG Score	-0.000	0.000
	(0.001)	(0.001)
Environmental Pillar		
Social Pillar		
Governance Pillar		
Size (log Mkt Cap)	-0.028*	-0.013
	(0.017)	(0.011)
Market-to-Book	0.007	-0.001
	(0.005)	(0.003)
ROA	-0.233	-0.139
	(0.243)	(0.138)
Leverage	-0.490***	-0.295***
	(0.120)	(0.075)
CFO (scaled)	-0.029	-1.313***
	(0.256)	(0.154)
log(1+Firm Age)	-0.035*	-0.005

	(0.019)	(0.011)
Analyst Coverage	-0.003	-0.004*
	(0.003)	(0.002)
Big4 Auditor	-0.042	-0.022
	(0.039)	(0.024)
Year FE	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes
Observations	1,545	1,413
R-squared	0.269	0.143
Adj. R-squared	0.245	0.115

Panel B. Environmental Pillar

Dependent variable: |DA_KM| / Total REM		
ESG Score		
Environmental Pillar	-0.002*	0.000
	(0.001)	(0.001)
Social Pillar		
Governance Pillar		
Size (log Mkt Cap)	-0.022	-0.013
	(0.016)	(0.011)
Market-to-Book	0.006	-0.001
	(0.005)	(0.003)
ROA	-0.250	-0.137
	(0.243)	(0.139)
Leverage	-0.487***	-0.294***
	(0.120)	(0.075)
CFO (scaled)	-0.053	-1.315***
	(0.256)	(0.153)
log(1+Firm Age)	-0.028	-0.006

	(0.019)	(0.011)
Analyst Coverage	-0.002	-0.004*
	(0.003)	(0.002)
Big4 Auditor	-0.041	-0.022
	(0.038)	(0.024)
Year FE	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes
Observations	1,545	1,413
R-squared	0.270	0.143
Adj. R-squared	0.247	0.115

Panel C. Social Pillar

Dependent variable: |DA_KM| / Total REM		
ESG Score		
Environmental Pillar		
Social Pillar	0.001	-0.001
	(0.001)	(0.001)
Governance Pillar		
Size (log Mkt Cap)	-0.034**	-0.009
	(0.017)	(0.011)
Market-to-Book	0.008*	-0.002
	(0.005)	(0.003)
ROA	-0.236	-0.144
	(0.243)	(0.138)
Leverage	-0.495***	-0.291***
	(0.120)	(0.075)
CFO (scaled)	-0.018	-1.319***
	(0.256)	(0.153)
log(1+Firm Age)	-0.036*	-0.005

	(0.019)	(0.011)
Analyst Coverage	-0.003	-0.003
	(0.003)	(0.002)
Big4 Auditor	-0.046	-0.022
	(0.039)	(0.024)
Year FE	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes
Observations	1,545	1,413
R-squared	0.269	0.144
Adj. R-squared	0.246	0.116

Panel D. Governance Pillar

Dependent variable: |DA_KM| / Total REM		
ESG Score		
Environmental Pillar		
Social Pillar		
Governance Pillar	0.001	0.001*
	(0.001)	(0.000)
Size (log Mkt Cap)	-0.031*	-0.014
	(0.016)	(0.010)
Market-to-Book	0.007	-0.001
	(0.005)	(0.002)
ROA	-0.234	-0.139
	(0.243)	(0.138)
Leverage	-0.501***	-0.308***
	(0.120)	(0.076)
CFO (scaled)	-0.008	-1.284***
	(0.256)	(0.154)
log(1+Firm Age)	-0.033*	-0.004

	(0.019)	(0.011)
Analyst Coverage	-0.004	-0.004**
	(0.003)	(0.002)
Big4 Auditor	-0.046	-0.025
	(0.039)	(0.024)
Year FE	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes
Observations	1,545	1,413
R-squared	0.269	0.145
Adj. R-squared	0.246	0.116

*Note: Each panel replaces the aggregate ESG score with a single pillar score (Environmental, Social, or Governance), holding the rest of the specification identical to Table 8. Note: Robust standard errors clustered at the firm and year levels in parentheses. *** p<0.01, ** p<0.05, * p<0.10.*

Appendix C. Individual REM Components

Variable	**Abn CFO**	**Abn PROD**	**Abn DISEXP**
Panel A. Aggregate ESG, no substitution control			
Dependent variable: Abn CFO / PROD / DISEXP			
ESG Score	-0.000	-0.000	-0.000
	(0.000)	(0.001)	(0.000)
Environmental Pillar			
|DA_KM| (AEM)			
Size (log Mkt Cap)	0.002**	-0.013	-0.000
	(0.001)	(0.010)	(0.003)
Market-to-Book	-0.000	0.002	0.003***
	(0.000)	(0.002)	(0.001)

ROA	0.006	-0.130	0.011
	(0.012)	(0.130)	(0.043)
Leverage	-0.004	-0.312***	-0.011
	(0.007)	(0.071)	(0.023)
CFO (scaled)	0.881***	-0.282**	0.140***
	(0.014)	(0.144)	(0.048)
log(1+Firm Age)	-0.001	-0.001	0.006
	(0.001)	(0.011)	(0.003)
Analyst Coverage	-0.001***	-0.002	0.002***
	(0.000)	(0.002)	(0.001)
Big4 Auditor	-0.002	-0.011	0.011
	(0.002)	(0.023)	(0.007)
Year FE	Yes	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes	Yes
Observations	1,413	1,413	1,413
R-squared	0.815	0.075	0.099
Adj. R-squared	0.809	0.045	0.069
Panel B. Aggregate ESG + |DA_KM| substitution control			
Dependent variable: Abn CFO / PROD / DISEXP			
ESG Score	-0.000	0.001	-0.001**
	(0.000)	(0.001)	(0.000)
Environmental Pillar			
|DA_KM| (AEM)	0.001	0.062***	0.001
	(0.002)	(0.018)	(0.006)
Size (log Mkt Cap)	0.003**	-0.016	0.005
	(0.001)	(0.012)	(0.004)
Market-to-Book	-0.001***	0.001	0.004***
	(0.000)	(0.003)	(0.001)
ROA	-0.004	-0.131	0.020
	(0.014)	(0.153)	(0.052)

Leverage	0.002	-0.249***	0.010
	(0.008)	(0.082)	(0.028)
CFO (scaled)	0.923***	-0.242	0.180***
	(0.016)	(0.171)	(0.058)
log(1+Firm Age)	-0.001	0.005	0.004
	(0.001)	(0.012)	(0.004)
Analyst Coverage	-0.001***	-0.000	0.002**
	(0.000)	(0.002)	(0.001)
Big4 Auditor	-0.002	-0.065**	0.007
	(0.002)	(0.026)	(0.009)
Year FE	Yes	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes	Yes
Observations	1,024	1,024	1,024
R-squared	0.835	0.116	0.137
Adj. R-squared	0.828	0.076	0.097

Panel C. Environmental Pillar + |DA_KM| substitution control

Dependent variable: Abn CFO / PROD / DISEXP			
ESG Score			
Environmental Pillar	-0.000	0.001**	-0.000
	(0.000)	(0.001)	(0.000)
|DA_KM| (AEM)	0.001	0.064***	0.001
	(0.002)	(0.018)	(0.006)
Size (log Mkt Cap)	0.003**	-0.019	0.004
	(0.001)	(0.012)	(0.004)
Market-to-Book	-0.001***	0.002	0.004***
	(0.000)	(0.003)	(0.001)
ROA	-0.004	-0.121	0.020
	(0.014)	(0.153)	(0.052)
Leverage	0.001	-0.250***	0.005
	(0.008)	(0.081)	(0.027)

CFO (scaled)	0.924***	-0.227	0.186***
	(0.016)	(0.170)	(0.058)
log(1+Firm Age)	-0.000	-0.000	0.005
	(0.001)	(0.013)	(0.004)
Analyst Coverage	-0.001***	-0.000	0.002**
	(0.000)	(0.002)	(0.001)
Big4 Auditor	-0.002	-0.063**	0.006
	(0.002)	(0.026)	(0.009)
Year FE	Yes	Yes	Yes
Industry (NAICS Subsector) FE	Yes	Yes	Yes
Observations	1,024	1,024	1,024
R-squared	0.836	0.120	0.135
Adj. R-squared	0.828	0.080	0.096

*Note: Decomposes the Total REM regression in Table 8 into the three operational channels of Roychowdhury (2006): abnormal cash flow from operations (Abn CFO), abnormal production costs (Abn PROD), and abnormal discretionary expenses (Abn DISEXP). Panel A reports the aggregate-ESG specification; Panel B adds |DA_KM| as a substitution control; Panel C replaces aggregate ESG with the Environmental Pillar score. Note: Robust standard errors clustered at the firm and year levels in parentheses. *** p<0.01, ** p<0.05, * p<0.10.*

Appendix D. Additional Causal Tests

Variable	**AEM equation**	**REM equation**
Panel A. Joint-Substitution OLS (Aggregate ESG)		
Dependent variable: |DA_KM| / Total REM		
ESG Score	-0.003*	0.001
	(0.0017)	(0.001)
Total REM	0.156***	
	(0.054)	
|DA_KM| (AEM)		0.055***
		(0.019)

Treat × Post (IV)		
Placebo Treat × Post		
Size (log Mkt Cap)	-0.012	-0.022*
	(0.021)	(0.013)
Market-to-Book	0.005	-0.003
	(0.006)	(0.003)
ROA	-0.351	-0.133
	(0.273)	(0.162)
Leverage	-0.447***	-0.259***
	(0.146)	(0.087)
CFO (scaled)	0.236	-1.351***
	(0.314)	(0.181)
log(1+Firm Age)	-0.035	0.001
	(0.022)	(0.013)
Analyst Coverage	-0.000	-0.001
	(0.004)	(0.002)
Big4 Auditor	-0.058	-0.073***
	(0.046)	(0.027)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	1,024	1,024
R-squared	0.313	0.189
Adj. R-squared	0.282	0.151
Panel B. Reduced-Form DiD / Intent-to-Treat		
Dependent variable: |DA_KM| / Total REM		
ESG Score		
Total REM		

|DA_KM| (AEM)		
Treat × Post (IV)	-0.032	-0.081
	(0.071)	(0.059)
Placebo Treat × Post		
Size (log Mkt Cap)	0.007	0.058**
	(0.096)	(0.023)
Market-to-Book	0.015	-0.006
	(0.010)	(0.004)
ROA	-0.025	0.161
	(0.260)	(0.105)
Leverage	-0.073	0.295
	(0.237)	(0.201)
CFO (scaled)	0.118	-0.506**
	(0.738)	(0.188)
log(1+Firm Age)	-0.225	-0.210*
	(0.237)	(0.108)
Analyst Coverage	-0.006	-0.002
	(0.007)	(0.003)
Big4 Auditor	0.076	0.062
	(0.044)	(0.037)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	986	847
R-squared	0.368	0.779
Adj. R-squared	0.265	0.740
Panel C. Placebo Test (2013 false treatment in 2010-2016 subsample)		
Dependent variable: ESG Score / |DA_KM|		
ESG Score		

Total REM		
|DA_KM| (AEM)		
Treat × Post (IV)		
Placebo Treat × Post	2.317	
	(1.769)	
Size (log Mkt Cap)	-0.660	0.051
	(0.910)	(0.033)
Market-to-Book	0.247	0.013
	(0.296)	(0.008)
ROA	1.748	0.404
	(4.041)	(0.235)
Leverage	4.216	0.032
	(6.419)	(0.228)
CFO (scaled)	-4.489	-1.555
	(5.605)	(0.814)
log(1+Firm Age)	20.882**	0.090
	(6.989)	(0.243)
Analyst Coverage	0.159	-0.010
	(0.108)	(0.011)
Big4 Auditor	-0.311	0.035
	(1.229)	(0.078)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	483	189
R-squared	0.933	0.440
Adj. R-squared	0.908	0.084

*Note: Panel A reports the joint-substitution OLS specification, in which the AEM equation includes Total REM as a cross-control and the REM equation includes |DA_KM|. Panel B reports the reduced-form difference-in-differences (intent-to-treat) coefficient on Treat × Post. Panel C is a placebo test on the pre-NFRD subsample (2010-2016) using 2013 as a false treatment date; insignificant or absorbed placebo coefficients support the absence of differential pre-trends. Note: Robust standard errors clustered at the firm and year levels in parentheses. *** p<0.01, ** p<0.05, * p<0.10.*

Appendix E Usage of AI Tools in Thesis Writing

Several AI assistants were consulted during the preparation of this thesis to support coding work, language refinement, and the formatting of tabular output. The tools employed were ChatGPT (OpenAI) Gemini (Google) and Grammarly (Superhuman Platform Inc.), each used for a distinct set of tasks that are described below. The use of these tools did not extend to the formulation of research questions, the design of the empirical strategy, the interpretation of results, or the drafting of original analytical arguments. All such intellectual content remains the work of the authors. Every suggestion produced by an AI tool was treated as a draft input requiring verification, and the procedures used to validate that input are summarised here for transparency.

In the empirical work conducted in R, the assistants were used principally to interpret error messages, to suggest alternative implementations of statistical procedures, and to clarify the syntax of less familiar functions. Suggested code was not adopted on trust: every function call was checked against the official documentation of the relevant package, and the numerical output of every procedure was validated by manual cross-checks, by comparison with worked examples from the methodological literature, and where possible by alternative implementations producing the same quantity.

For the written text, the assistants were used to refine grammar, sentence structure, and academic phrasing, and to identify passages where the exposition could be tightened or clarified. The substantive content of every paragraph, including the theoretical claims, the empirical interpretations, and the engagement with prior literature, was authored by the author of this thesis. AI assistance was confined to surface-level linguistic refinement. Each suggested edit was reviewed to ensure that the original meaning was preserved and that no technical detail was lost or distorted in the process. The formatting of a small number of tables was assisted by AI tools working from screenshots of R output, and every such table was checked against the underlying numerical results to confirm that the figures and labels were correctly transcribed. No part of this thesis was generated by an AI tool without subsequent human verification.

Appendix F. Industry-Cell Minimum-Size Sensitivity

The cross-sectional estimation of the Kothari et al. (2005) discretionary-accruals model at the NAICS subsector-year level requires a minimum number of firm-years per cell to produce well-identified normal-accrual benchmarks. The main analysis imposes a threshold of ten observations per cell, following common practice in the European earnings-management literature. Thinly populated cells with only ten to fourteen firms may generate noisier benchmarks against which discretionary residuals are measured, and a tightening of the threshold should therefore reduce the noise component of the |DA_KM| proxy. Appendix G reports the sample size and the mean absolute value of discretionary accruals at three threshold choices: the baseline ten observations per cell, a stricter fifteen, and the most demanding twenty. As the threshold tightens, the estimable sample contracts from 7,233 firm-years to 5,343, and the mean of |DA_KM| falls from 0.355 to 0.235, a decline consistent with the removal of noisy small-cell residuals that mechanically inflate the proxy at the loosest threshold. Importantly, the sign and magnitude of this attenuation are in the dependent variable rather than the explanatory variable of interest. The headline H1 conclusion of a statistically null ESG coefficient is unaffected by the noise content of the |DA_KM| proxy, since the noise component is additive to the residual rather than correlated with the right-hand-side regressor of interest. The sensitivity analysis therefore confirms that the headline inference reported in Table 8 of Section 5 does not rest on the particular minimum-cell choice used in the main implementation.

Appendix G. Industry-Cell Minimum-N Sensitivity (Kothari Model)

Min_Cell_Size	N_FirmYears_Estimated	Mean_Abs_DA
10	7,233	0.3552
15	6,199	0.2998
20	5,343	0.2351

Notes: Each row reports the estimable sample of firm-years and the mean absolute value of discretionary accruals (Kothari et al., 2005) when the cross-sectional estimation of the discretionary-accruals model is restricted to NAICS subsector-year cells containing at least the indicated number of firm-years. Row 1 corresponds to the baseline implementation used throughout the analysis. The progressive decline in the mean |DA_KM| as the threshold tightens is consistent with the removal of noisy small-cell residuals, and confirms that the headline H1 ESG coefficient reported in Table 8 of Section 5 is not driven by noise in the dependent variable.

Reference List

Badertscher, B. A. 2011, "Overvaluation and the Choice of Alternative Earnings Management Mechanisms", The Accounting Review, vol. 86, no. 5, pp. 1491-1518.

Barton, J. & Simko, P. J. 2002, "The Balance Sheet as an Earnings Management Constraint", The Accounting Review, vol. 77, no. s-1, pp. 1-27.

Bozzolan, S., Fabrizi, M., Mallin, C. A. & Michelon, G. 2015, "Corporate Social Responsibility and Earnings Quality: International Evidence", The International Journal of Accounting, vol. 50, no. 4, pp. 361-396.

Carroll, A. B. 1979, "A Three-Dimensional Conceptual Model of Corporate Performance", Academy of Management Review, vol. 4, no. 4, pp. 497-505.

Chen, Y. C., Hung, M. & Wang, Y. 2018, "The Effect of Mandatory CSR Disclosure on Firm Profitability and Social Externalities: Evidence from China", Journal of Accounting and Economics, vol. 65, no. 1, pp. 169-190.

Chih, H. L., Shen, C. H. & Kang, F. C. 2008, "Corporate Social Responsibility, Investor Protection, and Earnings Management: Some International Evidence", Journal of Business Ethics, vol. 79, no. 1-2, pp. 179-198.

Christensen, H. B., Hail, L. & Leuz, C. 2021, "Mandatory CSR and Sustainability Reporting: Economic Analysis and Literature Review", Review of Accounting Studies, vol. 26, no. 3, pp. 1176-1248.

Cohen, D. A., Dey, A. & Lys, T. Z. 2008, "Real and Accrual-Based Earnings Management in the Pre- and Post-Sarbanes-Oxley Periods", The Accounting Review, vol. 83, no. 3, pp. 757-787.

Cohen, D. A. & Zarowin, P. 2010, "Accrual-Based and Real Earnings Management Activities around Seasoned Equity Offerings", Journal of Accounting and Economics, vol. 50, no. 1, pp. 2-19.

Dechow, P. M., Hutton, A. P., Kim, J. H. & Sloan, R. G. 2012, "Detecting Earnings Management: A New Approach", Journal of Accounting Research, vol. 50, no. 2, pp. 275-334.

Dechow, P. M. & Skinner, D. J. 2000, "Earnings Management: Reconciling the Views of Accounting Academics, Practitioners, and Regulators", Accounting Horizons, vol. 14, no. 2, pp. 235-250.

Dhaliwal, D. S., Li, O. Z., Tsang, A. & Yang, Y. G. 2011, "Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting", The Accounting Review, vol. 86, no. 1, pp. 59-100.

Elrazaz, T. & Aljifri, K. 2026, "ESG and Earnings Management Nexus in the GCC: Evidence from a Quantile-on-Quantile Regression Approach", International Review of Economics and Finance, vol. 105, p. 104818.

Fiechter, P., Hitz, J. M. & Lehmann, N. 2022, "Real Effects of a Widespread CSR Reporting Mandate: Evidence from the European Union's NFRD", Journal of Accounting Research, vol. 60, no. 4, pp. 1499-1549.

Fombrun, C. & Shanley, M. 1990, "What's in a Name? Reputation Building and Corporate Strategy", Academy of Management Journal, vol. 33, no. 2, pp. 233-258.

Francis, J. R., Maydew, E. L. & Sparks, H. C. 1999, "The Role of Big 6 Auditors in the Credible Reporting of Accruals", Auditing: A Journal of Practice & Theory, vol. 18, no. 2, pp. 17-34.

Freeman, R. E. 1998, Strategic Management: A Stakeholder Approach, Pitman, Boston.

Garriga, E. & Melé, D. 2004, "Corporate Social Responsibility Theories: Mapping the Territory", Journal of Business Ethics, vol. 53, no. 1-2, pp. 51-71.

Gavana, G., Gottardo, P. & Moisello, A. M. 2022, "Related Party Transactions and Earnings Management: The Moderating Effect of ESG Performance", Sustainability, vol. 14, no. 10, p. 5823.

Godfrey, P. C. 2005, "The Relationship between Corporate Philanthropy and Shareholder Wealth: A Risk Management Perspective", Academy of Management Review, vol. 30, no. 4, pp. 777-798.

Gow, I. D., Ormazabal, G. & Taylor, D. J. 2010, "Correcting for Cross-Sectional and Time-Series Dependence in Accounting Research", The Accounting Review, vol. 85, no. 2, pp. 483-512.

Graham, J. R., Harvey, C. R. & Rajgopal, S. 2005, "The Economic Implications of Corporate Financial Reporting", Journal of Accounting and Economics, vol. 40, no. 1-3, pp. 3-73.

Gras-Gil, E., Palacios-Manzano, M. & Hernández-Fernández, J. 2016, "Investigating the Relationship between Corporate Social Responsibility and Earnings Management: Evidence from Spain", BRQ Business Research Quarterly, vol. 19, no. 4, pp. 289-299.

Grewal, J., Riedl, E. J. & Serafeim, G. 2019, "Market Reaction to Mandatory Nonfinancial Disclosure", Management Science, vol. 65, no. 7, pp. 3061-3084.

Hausman, J. A. 1978, "Specification Tests in Econometrics", Econometrica, vol. 46, no. 6, pp. 1251-1271.

He, F., Du, H. & Yu, B. 2022, "Corporate ESG Performance and Manager Misconduct: Evidence from China", International Review of Financial Analysis, vol. 82, p. 102201.

Healy, P. M. & Wahlen, J. M. 1999, "A Review of the Earnings Management Literature and Its Implications for Standard Setting", Accounting Horizons, vol. 13, no. 4, pp. 365-383.

Jensen, M. C. & Meckling, W. H. 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", Journal of Financial Economics, vol. 3, no. 4, pp. 305-360.

Jones, T. M. 1995, "Instrumental Stakeholder Theory: A Synthesis of Ethics and Economics", Academy of Management Review, vol. 20, no. 2, pp. 404-437.

Kasznik, R. 1999, "On the Association between Voluntary Disclosure and Earnings Management", Journal of Accounting Research, vol. 37, no. 1, pp. 57-81.

Kim, Y., Park, M. S. & Wier, B. 2012, "Is Earnings Quality Associated with Corporate Social Responsibility?", The Accounting Review, vol. 87, no. 3, pp. 761-796.

Kleibergen, F. & Paap, R. 2006, "Generalized Reduced Rank Tests Using the Singular Value Decomposition", Journal of Econometrics, vol. 133, no. 1, pp. 97-126.

Kothari, S. P., Leone, A. J. & Wasley, C. E. 2005, "Performance Matched Discretionary Accrual Measures", Journal of Accounting and Economics, vol. 39, no. 1, pp. 163-197.

Larcker, D. F. & Rusticus, T. O. 2010, "On the Use of Instrumental Variables in Accounting Research", Journal of Accounting and Economics, vol. 49, no. 3, pp. 186-205.

Liang, H. & Renneboog, L. 2017, "On the Foundations of Corporate Social Responsibility", The Journal of Finance, vol. 72, no. 2, pp. 853-909.

LSEG 2024, Environmental, Social and Governance Scores from LSEG (Methodology document), London Stock Exchange Group, viewed 20 June 2026, https://www.lseg.com/content/dam/data-analytics/en_us/documents/methodology/lseg-esg-scores-methodology.pdf.

Moharram, A. H., Alahdal, W. M., Adnan, S. M., Hussainey, K. & Hashim, H. A. 2026, "The Effect of ESG Disclosure on Firm Performance: Does Earnings Management Matter?", Business Strategy and the Environment.

Ormiston, M. E. & Wong, E. M. 2013, "License to Ill: The Effects of Corporate Social Responsibility and CEO Moral Identity on Corporate Social Irresponsibility", Personnel Psychology, vol. 66, no. 4, pp. 861-893.

Petersen, M. A. 2009, "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches", The Review of Financial Studies, vol. 22, no. 1, pp. 435-480.

Petrovits, C. M. 2006, "Corporate-Sponsored Foundations and Earnings Management", The Accounting Review, vol. 81, no. 2, pp. 335-361.

Prior, D., Surroca, J. & Tribó, J. A. 2008, "Are Socially Responsible Managers Really Ethical? Exploring the Relationship Between Earnings Management and Corporate Social Responsibility", Corporate Governance: An International Review, vol. 16, no. 3, pp. 160-177.

Ricapito, F. P. 2024, "Earnings Management and ESG Performance: Empirical Evidence from Italian Context", Corporate Ownership and Control, vol. 21, no. 2, p. 86.

Roberts, M. R. & Whited, T. M. 2013, "Endogeneity in Empirical Corporate Finance", in Handbook of the Economics of Finance, eds. G. M. Constantinides, R. M. Stulz & M. Harris, Elsevier B.V., vol. 2A, pp. 493-572.

Roychowdhury, S. 2006, "Earnings Management through Real Activities Manipulation", Journal of Accounting and Economics, vol. 42, no. 3, pp. 335-370.

Stock, J. H. & Yogo, M. 2005, "Testing for Weak Instruments in Linear IV Regression", in Cambridge University Press eBooks, Cambridge University Press, pp. 80-108.

Sun, W., Chen, S., Jiao, Y. & Feng, X. 2024, "How Does ESG Constrain Corporate Earnings Management? Evidence from China", Finance Research Letters, vol. 61, p. 104983.

Velte, P. 2021, "Environmental Performance, Carbon Performance and Earnings Management: Empirical Evidence for the European Capital Market", Corporate Social Responsibility and Environmental Management, vol. 28, no. 1, pp. 42-53.

Verschoor, C. C. 1998, "A Study of the Link between a Corporation's Financial Performance and Its Commitment to Ethics", Journal of Business Ethics, vol. 17, no. 13, pp. 1509-1516.

Wadhaad, A. S., Chebbi, K. & Ammer, M. A. 2023, "Unveiling the Connection among ESG, Earnings Management, and Financial Distress: Insights from an Emerging Market", Sustainability, vol. 15, no. 16, p. 12348.

Watts, R. L. & Zimmerman, J. L. 1990, "Positive Accounting Theory: A Ten Year Perspective", The Accounting Review, vol. 65, no. 1, pp. 131-156.

Wooldridge, J. M. & Li, T. 2002, "Econometric Analysis of Cross Section and Panel Data", Journal of Economic Literature, vol. 40, no. 4, pp. 1239-1241.

Wu, D.-M. 1973, "Alternative Tests of Independence between Stochastic Regressors and Disturbances", Econometrica, vol. 41, no. 4, pp. 733-750.

Xue, C., Li, R., Xue, S. & Zhang, X. 2025, "Mandatory versus Voluntary: The Real Effect of ESG Disclosures on Corporate Earnings Management", Journal of International Money and Finance, vol. 154, p. 103323.

Zang, A. Y. 2012, "Evidence on the Trade-Off between Real Activities Manipulation and Accrual-Based Earnings Management", The Accounting Review, vol. 87, no. 2, pp. 675-703.