Effects of ESG Uncertainty on Pension Funds' Sustainable Investment Disclosure

EPA2942: Master Thesis

Max Vos

Delft University of Technology

TU Delft

Effects of ESG Uncertainty on Pension Funds' Sustainable Investment Disclosure

by

Max Vos

Max

Vos

First Supervisor:	E. Schröder
Second Supervisor and Chair:	L. Kamp
Advisor:	A. Ralcheva
Company Supervisor:	S. van den Beld
Project Duration:	November, 2025 - March, 2026
Faculty:	Technology, Policy and Management, Delft

Cover: AI-generated cover design created using ChatGPT (GPT-5.2). The background image is derived from the original frontispiece of *Leviathan* (1651) by Thomas Hobbes (engraving by Abraham Bosse). The reference reflects the idea that, when expectations are uncertain, supervision can act as a coordinating force that stabilises disclosure and transparency, a role examined in this thesis in the context of pension fund sustainability reporting.

TU Delft

Preface

This thesis looks at how debates and uncertainty around ESG affect how pension funds talk about sustainable investing. As rules change and political discussions grow, financial institutions often face unclear or disputed signals. Knowing how this uncertainty shapes what organisations share is important for financial governance, for designing policy tools that use information to guide behaviour, and especially for De Nederlandsche Bank's pension supervision.

This topic fits well with the Engineering and Policy Analysis (EPA) programme. It looks at how financial institutions, political actors, and rules interact, focusing on decisions made under uncertainty. By studying how debates and uncertainty affect what organisations share, this thesis adds to EPA's focus on policy design, the use of information in governance, and how strong policy tools are in challenging situations.

I completed this thesis during my graduation internship at De Nederlandsche Bank (DNB), in the department that oversees pension supervision. This setting gave the research a practical focus: annual reports are not just for public accountability, but also provide accessible information for risk-based supervision. Working within supervision helped me see what disclosure can and cannot reveal, and how changes in reporting matter when policy signals are unclear. I also appreciated the opportunity to discuss my work with authors of antecedents at DNB and to learn from their experience.

I chose this topic because I am interested in sustainable finance and how institutions make decisions, and I wanted to work in areas where I had less experience. This research meant using methods from Natural Language Processing and Econometrics. Turning annual reports into numbers and connecting them to a monthly uncertainty index was sometimes difficult, especially when trying to keep the methods both rigorous and easy to understand. Still, I am proud of developing a measurement approach that is both scalable and closely linked to a real supervisory issue.

This thesis is for researchers and professionals in pension supervision, sustainable finance, and financial regulation. It explains how policy tools based on disclosure work in practice when the environment is uncertain. By looking at how policy uncertainty affects what is reported about sustainable investing, the thesis shows that narrative reporting can be a simple way to monitor funds, especially smaller ones with less frequent supervision. It also points out the strengths and limits of relying on voluntary narrative disclosure during debates about ESG policies.

I want to thank Sven van den Beld and De Nederlandsche Bank for letting me do this research at their organisation and for providing access to their data. I am also grateful to my graduation committee members, Aleksandrina, Enno, and Linda, for their guidance and helpful feedback during the research. Finally, I thank my friends and family for their support and patience while I was writing.

I hope you enjoy reading this thesis.

Max Vos
Amsterdam, March 2026

Executive Summary

Pension funds play an important role in the sustainability transition. They invest for the long term and manage large parts of the economy. Many funds report on how they address environmental, social, and governance (ESG) risks and how they invest sustainably. At the same time, ESG has become more politically debated and less predictable. Expectations about what counts as “good” ESG behaviour can change quickly, and strong public claims about sustainability may bring reputational or legal risks. Because the day-to-day investment decisions of pension funds are difficult for outsiders to observe, annual reports are an important source of information about how funds approach sustainable investing (SI). However, reporting is not purely descriptive: organisations can adjust what they communicate when the environment around them becomes more uncertain.

This thesis examines how increases in ESG-related debate and uncertainty in public discourse affect the way Dutch pension funds report on sustainable investing. The focus is on whether uncertainty changes how prominently funds discuss sustainability and how broadly they describe their SI strategies.

To answer this question, the study combines two sources of information. First, it uses a monthly ESG Uncertainty Index (ESGUI) for the Netherlands. This index captures the amount of ESG-related debate and uncertainty language in news coverage and therefore reflects how contested and uncertain the context of ESG discussions is in public discourse. Second, the thesis constructs a new dataset of annual reports from 124 Dutch pension funds between 2018 and 2024.

Using modern text analysis methods, the annual reports are translated into five indicators of SI disclosure. These indicators measure how much funds discuss sustainable investing (Intensity), how many topics they cover (Spectrum), how concrete their statements are (Specificity), how many types of SI strategies they mention (Variety), and how broadly SI is described across asset classes (Scope). The analysis then examines whether these indicators change when ESG uncertainty is higher in the period leading up to the publication of each report.

The results suggest a clear but modest pattern. When ESG uncertainty is higher, pension funds tend to devote slightly less attention to sustainable investing in their annual reports and discuss it across fewer topics, strategies, and asset classes. In other words, SI reporting becomes somewhat narrower. At the same time, there is little evidence that the remaining language becomes less concrete. This indicates that funds mainly respond by reducing the breadth of their sustainability narrative rather than by switching to vaguer language.

These effects should be interpreted carefully. The estimated changes in reporting are relatively small and reflect communication choices rather than direct changes in investment behaviour. The strongest evidence for this contraction pattern appears in the measures that capture how widely sustainability is discussed (Variety and Scope). Other indicators are less robust and should therefore be interpreted more cautiously.

The results also show some differences between funds. Smaller pension funds tend to adjust their reporting somewhat more when ESG debates intensify, possibly because they have fewer resources to manage communication or face greater reputational risks. For funds with higher carbon exposure, the most noticeable adjustment appears in the range of sustainability strategies they describe.

Beyond the empirical results, the thesis also contributes methodologically by showing how text analysis and modern language models can be used to measure different aspects of sustainability disclosure. These tools make it possible to analyse large collections of reports and track how reporting patterns evolve over time.

The findings are particularly relevant for financial supervision, like De Nederlandsche Bank (DNB). Sustainability reporting is an important source of information for supervisors, but this thesis shows that the amount and scope of such reporting may change when ESG debates become more uncertain.

This means that reductions in disclosure do not necessarily imply changes in investment behaviour. At the same time, systematic text analysis could help supervisors monitor how sustainability reporting evolves across the sector and identify when reporting becomes narrower or less prominent. Additionally, monitoring the effects of DNB's own policies can be done automatically using these methods.

Overall, the thesis suggests that sustainability disclosure is not only a tool for transparency but can also function as a strategic communication instrument. When ESG debates intensify and expectations become less clear, pension funds may respond by narrowing their sustainability narrative. For supervisors and stakeholders, this stresses the importance of monitoring how sustainability reporting evolves over time, especially during periods when public debate around ESG becomes more uncertain.

Contents

1 Introduction

Pension funds are some of the world's largest asset owners, which gives them a special role in the financial sector's sustainability transition (Fabian et al., 2025). Their large assets and long-term outlook allow them to influence markets and direct investments toward more sustainable activities (McDonnell, 2024). They do this mainly in two ways. First, many funds include financially important environmental, social, and governance (ESG) factors in their standard risk and return processes. This is part of their long-term fiduciary duty (Edmans, 2023). Second, funds use Socially Responsible Investment (SRI; in Dutch: maatschappelijk verantwoord beleggen, MVB) policies. These policies reflect the values of members or society, often by excluding certain investments or focusing on specific themes (Bauer, Ruof, et al., 2021). In general, there are five main sustainable investment strategies: divestment, ESG integration, screening, public engagement, and private engagement (Bauer, Broeders, et al., 2023).

Pension funds invest in many different companies across the whole economy. Because of this, their long-term returns depend on the overall health of the economy, not just on the success of individual firms. This gives them a reason to endorse sustainable development, both by deciding where to invest their money and by using their influence as shareholders (Friede et al., 2015). These incentives are visible in public commitments such as pledges to reach net-zero emissions and to move away from fossil-fuel investments (Dohle, 2024). They are further strengthened by EU regulations, which require pension funds to consider environmental, social, and governance risks in their operations and investments (Pensions UK, 2024; PensionsEurope, 2023).

Over the past decade, public support for sustainable investing (SI) has generally increased (Bauer, Ruof, et al., 2021). At the same time, political backlash has also grown. This is especially clear in the United States, where leaders, including the president, have openly criticised ESG and sustainable investing. Political divisions have led to strong disagreements about its value (Cifrino, 2023; Ravina et al., 2025). As a result, states with different political leanings now take very different approaches: some require fund managers to consider ESG factors, while others ban them from doing so (Vasquez, 2023).

In Europe, these divisions are milder but still evident. Norway's sovereign wealth fund, the largest in the world, with assets of $2.1 trillion, recently paused planned ethical divestments following pressure from the United States (NOS, 2025; Fouche, 2025). In the Netherlands, Parliament passed a motion stating that pension funds should focus primarily on financial returns and avoid activist behaviour (Dohle, 2024). Illustrating these tensions, 80% of Dutch institutional investors report concerns about political pushback to ESG investing (BNP Paribas, 2024).

These developments make ESG and sustainability issues more uncertain for institutional investors. This uncertainty can arise from changing or unclear regulations, disagreements about what ESG investing means, shifts in political support, or uncertainty about how regulators, stakeholders, or the public will evaluate sustainability commitments. As a result, pension funds may find it difficult to anticipate which ESG practices will be expected, encouraged, or criticised in the near future, even if formal regulations do not change.

Such uncertainty may influence not only investment decisions but also how funds communicate about sustainable investing. When expectations around ESG practices are unclear, making specific commitments in public disclosures may create reputational or legal risks, while more general language allows for greater flexibility. In this way, uncertainty surrounding ESG debates can shape disclosure choices, even if underlying investment practices remain unchanged. While research on climate and policy uncertainty in financial markets is growing (Marotta et al., 2025; Gavriilidis, 2021), relatively little is known about how broader ESG-related uncertainty affects the sustainability disclosures of long-term investors such as pension funds. With this in mind, this thesis asks the following research question:

How do increases in ESG-related debate and uncertainty in public discourse affect the sustainable investing (SI) disclosures of Dutch pension funds?

This question matters because many sustainable investing (SI) practices by pension funds are not visible in portfolio holdings or balance sheets. ESG integration, company engagement, and setting or monitoring sustainability targets usually happen internally. As a result, annual reports and other disclosures are the main sources for beneficiaries, regulators, and other stakeholders to evaluate a fund's SI approach over time. In this thesis, I focus on communication instead of direct actions. The analysis cannot directly track changes in portfolios, engagement, or voting. Therefore, changes in disclosure should be seen as shifts in emphasis and framing, which may or may not reflect real changes in SI activity.

Disclosure is still important because it makes internal practices visible, helps close information gaps, clarifies expectations, and allows supervisors to assess whether funds meet regulatory and fiduciary duties (Verrecchia, 2001; Leuz et al., 2016). For supervisors, annual reports have two main roles: they inform beneficiaries and offer a practical, scalable way to monitor funds. Narrative ESG disclosure can act as an early signal of whether supervisory expectations are reflected in reported risk management, and can help identify cases where more detailed, fund-specific follow-up may be warranted, especially for smaller funds that have less direct contact with supervisors.

This becomes particularly relevant in environments where ESG-related expectations are contested or evolving. Uncertainty surrounding how sustainability commitments will be interpreted by regulators, stakeholders, or the public can increase the perceived risks associated with specific public statements, even if underlying practices remain unchanged. In such contexts, individually rational disclosure behaviour may shift toward preserving flexibility and limiting exposure to reputational or legal scrutiny. This can make transparency a collective good that is underprovided without coordination, a role that pension supervisors partly fulfil by setting expectations and stabilising disclosure practices.

An important premise of this thesis is that disclosure is a choice made under constraints. When expectations around ESG practices fluctuate or remain contested, pension funds may face a trade-off between demonstrating commitment and maintaining flexibility in their public communication. I therefore distinguish two plausible response modes that could, but need not, occur together. In a *contraction* response, funds reduce how much they say about SI and how many SI topics they cover, for instance, by omitting marginal topics or asset classes from the report. In a *hedging* response, funds keep talking about SI but use less concrete and less verifiable language, reducing the likelihood that they will be held to specific statements if standards or expectations change.

Based on disclosure and legitimacy theories, earlier research suggests that organisations may change what they communicate to protect their right to operate (Verrecchia, 2001; Leuz et al., 2016). These response types suggest two main ways disclosure can change: (i) the salience and breadth of SI reporting (how much and how widely funds discuss SI), and (ii) the concreteness and verifiability of SI statements (how specific the language is). At the same time, standardised reporting frameworks like the Task Force on Climate-related Financial Disclosures (TCFD) and the Sustainable Finance Disclosure Regulation (SFDR) can set a disclosure floor by requiring certain categories and metrics. This means that most changes may happen in the more flexible parts of annual reports, such as narrative emphasis, topic choice, and how funds describe their SI policies, rather than in the required sections. Whether contestation and uncertainty change disclosure even with these floors is important for designing effective disclosure frameworks.

Previous research shows that strong environmental performance and high-quality reporting often go together (Velte, 2023), but this connection can weaken or even reverse if disclosure is mainly used for impression management or greenwashing (Marquis et al., 2016; Gull et al., 2023; Cho, Laine, et al., 2015). With greater political scrutiny, some investors adopt a “quiet” ESG approach: they continue to consider sustainability risks internally while reducing visible ESG branding or public commitments (Ravina et al., 2025). These patterns support a focus on disclosure, but also caution against interpreting any single reporting feature as a direct reflection of underlying practice. In this thesis, I therefore use specificity as a measure of how concrete and verifiable the language is, rather than as a direct proxy for sustainable investment action.

Fund characteristics may also shape how pension funds adjust disclosure in more uncertain ESG-relevant environments. Larger funds and those with stronger governance structures tend to adopt more advanced SI strategies and disclose more extensively (Bauer, Bogman, et al., 2020; Bauer, Broeders, et al., 2023; De Nederlandsche Bank, 2018). Board composition, such as age and gender diversity, has also been shown to influence SI ambition and implementation (Agnese et al., 2024; Arayakarnkul et al., 2022; Al-Shaer et al., 2016). Funds with more carbon-intensive portfolios may face greater transition and greenwashing risks when sustainability-related expectations are contested (Aerts et al., 2009; Golka et al., 2025). For this reason, I examine whether the relationship between ESG-relevant uncertainty and SI disclosure varies with fund size and portfolio carbon intensity.

To answer these questions, I use a monthly, text-based ESG Sustainability Uncertainty Index (ESGUI) for the Netherlands (Ongan et al., 2025), along with text-based measures of SI disclosure for 124 Dutch pension funds from 2018 to 2024. ESGUI is based on how often ESG-related and uncertainty-related keywords appear in country reports, so it shows how important sustainability issues are when there is a lot of uncertainty in public discussions. This means it does not focus only on uncertainty about ESG policy, but instead captures ESG-related topics discussed in uncertain economic and policy situations. I use the index as a proxy for the wider uncertainty around ESG decision-making in public conversations. Since ESGUI looks specifically at ESG as a financial and governance idea, it fits this study better than broader measures like climate policy uncertainty indices (Marotta et al., 2025; Gavriilidis, 2021).

From an empirical perspective, it is challenging to identify effects because the ESG Uncertainty Index is national and only changes over time, while pension fund disclosure also changes with general trends and events. To better match disclosure to the information available when reports are finished, I use the fact that annual reports come out in different months. I link each report to a rolling pre-publication window of the ESG Uncertainty Index. By using pension fund and report-year fixed effects, the main relationships are identified for the most part from differences in exposure caused by publication timing within the same year.

For the dependent variables, I analyse annual reports and construct five complementary SI disclosure indicators: *Intensity* (how much SI is discussed), *Specificity* (how concrete and verifiable SI statements are), *Spectrum* (how many SI topics are covered), *Variety* (how many SI strategy categories are mentioned), and *Scope* (how many asset classes SI claims are framed as covering). Building on Bauer, Broeders, et al. (2023), I extend the measurement approach by using large language models (LLMs) to improve the classification of nuanced SI language and to simplify and improve the labelling process. I then estimate fixed-effects panel regressions to examine how variation in ESG-related contestation and uncertainty is associated with changes in these disclosure margins, and how the associations differ across funds.

The results show that when ESG-relevant expectations are more contested or uncertain, SI disclosure is modestly less prominent and covers slightly fewer topics, strategies, and asset classes. However, the concreteness of SI language does not change substantially. These effects are somewhat stronger for smaller funds and, in some cases, for funds with higher carbon exposure. This suggests that organisational capacity and concerns about legitimacy may influence how funds fine-tune their public reporting in uncertain ESG-relevant environments.

This study contributes to three main research areas. First, it expands the literature on political and policy uncertainty by examining how uncertainty in ESG-relevant decision environments affects pension fund disclosure, rather than focusing solely on firms. Second, it adds to sustainable finance research by offering evidence from a European, particularly Dutch, context where ESG integration is expected and political contestation is growing (De Nederlandsche Bank, 2018; Golka et al., 2025). Third, it makes a methodological contribution by building on Bauer, Broeders, et al. (2023) and using advanced text-based methods, including LLM-based classification, to measure different aspects of SI disclosure over time while reducing manual work.

The findings also have practical value. They indicate that pension fund sustainability disclosures largely reflect how information is presented in annual reports, rather than providing direct evidence of SI actions. In more uncertain ESG-relevant environments, voluntary narrative disclosures may become slightly less comprehensive, for instance, through the omission of marginal topics or asset classes, even if underlying practices remain unchanged. Although structured reporting templates set a basic standard, they may not fully prevent funds from adjusting how extensively sustainability practices are discussed in other parts of the reports. This has implications for supervision and cross-fund comparisons, particularly during periods of public debate around ESG. The thesis also offers a scalable approach to monitoring disclosures that can support proportional pension supervision. By breaking down narrative annual reports into different disclosure dimensions, such as salience, breadth, portfolio framing, and verifiability, this method helps track whether supervisory expectations and policy measures are reflected in reporting across the sector. This is especially useful for smaller funds, where supervisors have less direct contact and simple monitoring signals are most informative.

The rest of this thesis is organised as follows. Chapter 2 sets out the theory and hypotheses about how contestation and uncertainty might cause contraction or hedging in SI disclosure. Chapter 3 covers the data sources and variables. Chapter 4 explains how the text-based disclosure measures and ESGUI exposure mapping are built. Chapter 5 presents the empirical design, main results, analyses of moderation, and robustness tests.

2

Literature Review and Hypotheses

2.1. ESG-related Uncertainty and SI Disclosure

Political and policy uncertainty shapes how firms and investors act and communicate. When future rules, enforcement, or political priorities are unclear, organisations face higher regulatory and reputational risks. As a result, they may delay investments, take fewer risks, or change how they present their plans and commitments (Pastor et al., 2011; Hassan et al., 2019). For pension funds, which are closely watched by the public and supervisors, these communication choices matter even more.

Researchers often measure policy uncertainty with news and text-based indices that track how often policy and uncertainty-related words appear together (Baker et al., 2016). Newer measures focus on climate or environmental policy uncertainty, capturing unclear climate targets, sustainable finance rules, or environmental risk management (Gavriilidis, 2021; Marotta et al., 2025). These specific measures are especially important for long-term investors.

However, uncertainty around ESG-related investment and reporting decisions can come not only from unclear policies, but also from conflicting expectations, changing political support, or new interpretations of sustainability commitments. In this thesis, I use the ESG-based Sustainability Uncertainty Index (ESGUI) developed by Ongan et al. (2025) as a way to represent the broader ESG-related uncertainty in public discussions. ESGUI combines ESG-related attention and uncertainty-related language in country reports, showing times when ESG issues are both important and discussed in uncertain economic and policy situations. It does not focus only on uncertainty about ESG policy, but captures ESG-related decisions made during times of high uncertainty in public debate.

In the Netherlands, debates about climate and ESG policy have become more common over the past twenty years. After the Paris Agreement, climate policy uncertainty grew and became more negative, which suggests that people increasingly see climate policy as a source of transition costs and risks (UNFCCC, 2015; Marotta et al., 2025). Investors connect this situation to unstable policy signals, higher transition costs, and more regulatory scrutiny (Olasehinde-Williams et al., 2023). Earlier research shows that higher uncertainty leads to less investment and more cautious strategies, including in sustainability decisions (Gulen et al., 2016; Jens, 2017; Shaikh, 2022).

Many activities in sustainable investing (SI) are hard to see directly. ESG integration, engagement, and setting targets often happen inside the organisation. Because of this, disclosure is the main way beneficiaries, regulators, and other stakeholders can judge what a pension fund says it is doing. Reporting frameworks like TCFD, SFDR, and the EU taxonomy support this by encouraging more structured and comparable disclosure (TCFD, 2023; European Commission, 2024; Krueger et al., 2023). These frameworks also set a minimum standard by requiring certain categories and metrics. As a result, funds may adjust disclosure mostly in the more flexible parts of annual reports, such as which topics they highlight and how broadly they frame sustainability claims.

Disclosure is not always neutral. According to disclosure theory, managers decide what to share and how specific to be by balancing the benefits of transparency with legal, political, and competitive risks

(Verrecchia, 2001; Leuz et al., 2016). When the risk of being wrong is high, organisations may use cautious language and avoid making strong promises (Osovsky, 2015). If oversight and enforcement are weak, organisations can meet formal requirements with general statements instead of clear, verifiable information (Christensen et al., 2021). Anti-greenwashing efforts aim to close this gap, but enforcement is still developing and faces challenges (European Securities and Markets Authority, 2024; Muñoz et al., 2024).

Disclosure can change in more than one way. When expectations around ESG practices are more debated or less predictable, pension funds may face a trade-off between demonstrating commitment and maintaining flexibility in their public communication. This thesis identifies two plausible response modes, which do not always occur simultaneously.

First, funds may respond by **contracting**. In this approach, they make SI disclosure modestly less prominent and cover fewer SI topics, strategies, or portfolio segments, for instance, by omitting marginal topics or asset classes from the report. This may help limit exposure to political criticism or reduce the perceived risk that specific statements could later be challenged if standards or expectations evolve.

Second, funds may respond by **hedging**. Here, they continue to discuss SI but use less specific and less verifiable language. Instead of giving targets, timelines, KPIs, or clear coverage statements, reports may place greater emphasis on general principles or processes. This can preserve flexibility by lowering the likelihood that specific claims will be judged against changing supervisory, regulatory, or stakeholder expectations.

There is evidence that organisations under scrutiny manage reputational risk by using selective disclosure and 'coverage without commitment,' which fits both response modes (Marquis et al., 2016; Lyon et al., 2006; Delmas et al., 2011). Legitimacy theory also suggests that when expectations are debated, organisations may support broad principles but avoid commitments that can be monitored and enforced (Marquis et al., 2016; Gull et al., 2023; Luan, 2024; Cho, Laine, et al., 2015). Although strong environmental performance and strong reporting often go together (Velte, 2023), this link can weaken when disclosure becomes more symbolic.

Overall, the literature points to two likely responses when ESG-related expectations become more contested or less predictable. Funds may make SI communication less prominent and broad (contraction), use less concrete language in SI discussions (hedging), or do both. Since these responses do not always occur together, I treat them as separate hypotheses.

Importantly, literature shows that disclosure does not always map directly onto SI implementation. Some investors respond to ESG politicisation with a "quiet ESG" approach: they continue to manage sustainability risks internally while reducing visible ESG branding or public commitments (Ravina et al., 2025). This thesis therefore examines disclosure as the observable component of how pension funds respond to uncertainty in the ESG-relevant environment, rather than as a direct measure of underlying investment behaviour.

To connect the hypotheses to the measures used later, I separate outcomes into (i) salience and breadth, and (ii) concreteness and verifiability.

Salience and breadth are measured by Intensity (how much SI is discussed), Spectrum (how many SI topics are covered), Variety (how many SI strategy categories are mentioned), and Scope (how many asset classes SI claims cover). Concreteness and verifiability are measured by Specificity, which is the share of SI statements with checkable commitments such as targets, KPIs, timelines, or due dates.

H1a (Selective omission / contraction). When ESG-related issues are more salient in uncertain environments, Dutch pension funds reduce the *salience and breadth* of SI disclosure (lower Intensity and narrower Spectrum, Variety, and Scope).

H1b (Within-SI hedging / lower verifiability). When ESG-related issues are more salient in uncertain environments, Dutch pension funds' SI disclosures become *less concrete and less verifiable* (lower Specificity).

2.2. Visibility and Capacity

The relationship between uncertainty in the ESG-relevant environment and disclosure is unlikely to be uniform across pension funds. Differences in size, visibility, and organisational resources can shape both the incentives to disclose and the capacity to do so.

Larger and more visible organisations face more media and political attention (Healy et al., 2001). For pension funds, this visibility increases reputational risk because stakeholders are more likely to spot inconsistencies, gaps, or vague promises. This kind of scrutiny can encourage clearer reporting (Hahn et al., 2013; Brammer et al., 2006).

Larger funds usually have more resources, such as dedicated staff, established ESG processes, improved data systems, and more regular supervisory contact (Christensen et al., 2021; De Nederlandsche Bank, 2018). This may help them maintain broader and more consistent SI reporting even when sustainability-related expectations are contested or evolving (Aerts et al., 2009; Bauer, Broeders, et al., 2023). Smaller funds, by contrast, may simplify their reports, focus on fewer topics, or avoid making broad claims that are harder to substantiate under uncertain expectations.

Overall, these factors suggest that large funds have more stable reporting. While they may still manage reputational risk carefully, they usually have stronger reasons and more resources to keep their disclosure broad.

H2 (Visibility and capacity). The relationship between ESG-related debate and uncertainty and SI disclosure salience and breadth is weaker for larger funds than for smaller funds.

In the empirical analysis, fund size is treated as a largely time-invariant characteristic. H2 therefore concerns differences in the *slope* of the relationship across funds (large versus small), not the effect of a fund becoming larger over time.

2.3. Environmental Sensitivity

Funds with more carbon-intensive portfolios face different pressures. Their investments may seem at odds with sustainable finance goals, which increases legitimacy risks. This is similar to what firms in environmentally sensitive industries experience (Aerts et al., 2009; Cho, Laine, et al., 2015). When ESG expectations are unclear or debated, these funds may face closer scrutiny and a higher risk of being accused of inconsistency or greenwashing (Patten, 2002; Cho and Patten, 2007).

Since portfolios cannot be changed quickly (Bams et al., 2016), funds with higher carbon exposure may adjust disclosure more readily than underlying portfolio composition. For instance, they may narrow the scope of their claims, avoid controversial topics, or emphasise process and stewardship language that signals responsibility without making firm commitments (Delmas et al., 2011; Lyon et al., 2006; Bingler et al., 2022). Such responses may affect which aspects of sustainable investing are highlighted and how extensively SI is discussed in annual reports. As a result, funds with higher carbon exposure may show stronger disclosure adjustments when ESG debates become more uncertain.

H3 (Environmental sensitivity). The relationship between ESG-related debate and uncertainty and SI disclosure salience and breadth is stronger for funds with more carbon-intensive portfolios than for funds with lower carbon exposure.

In the empirical analysis, portfolio carbon intensity is treated as a largely time-invariant characteristic. H3 therefore concerns differences in the *slope* of the relationship across funds (higher versus lower carbon intensity), not short-run changes in carbon intensity within a fund.

2.4. Additional Determinants of SI Disclosure

Besides size and environmental exposure, other fund and board characteristics can also affect sustainable investing practices and ESG disclosure. The literature discusses these factors to give a broader context, but this thesis does not analyse them directly because of data limitations.

Studies find that larger and more established funds are more likely to have clear SI policies, interact with investee companies, and report on ESG topics in an organised way (Bauer, Bogman, et al., 2020; Bauer, Broeders, et al., 2023). Oversight methods that consider size and complexity, like proportional supervision by the Dutch central bank, can also improve the connection between fund characteristics and the quality of disclosure (De Nederlandsche Bank, 2018; Golka et al., 2025).

Board composition is also important. Studies in corporate governance and pension funds find that boards with more gender diversity and a broader mix of skills and backgrounds are more likely to focus on ESG risks and sustainability issues, and are linked to more detailed non-financial reporting (Al-Shaer et al., 2016; Arayakarnkul et al., 2022; Agnese et al., 2024). In European pension funds, these qualities are also connected to adopting responsible investment policies and including ESG factors in investment decisions (Bauer, Bogman, et al., 2020; Agnese et al., 2024).

These factors help explain why sustainable investing practices and disclosures vary across funds, and they may also affect how funds react to outside pressures like growing ESG debates or uncertainty in public discussions. Governance structures and board makeup can influence how organizations see reputational risks, regulatory demands, and stakeholder expectations. Because these features usually change slowly, this thesis focuses on mechanisms more closely tied to the main research question: how shifts in ESG debate and uncertainty show up in the prominence, range, and clarity of sustainable investing disclosures. Future studies could look at whether governance features like board composition or supervisory involvement affect how pension funds change their sustainability disclosures when ESG debates become more intense.

3

Data Sources and Variables

3.1. Dependent Variables: SI Measures

I study a sample of 124 Dutch pension funds over *2018-2024*. Annual reports were manually collected from pension funds' websites. Due to substantial variation in website architecture and document hosting practices across the 124 funds, automated web scraping was unlikely to yield comparable gains in efficiency or reliability.

Annual reports are the main disclosure source for three reasons. First, they are the most comprehensive public document in which pension funds explain their strategy, governance, and performance, including sustainable investing (SI) practices and priorities. Second, annual reports contain a large *voluntary narrative* component: beyond minimum required items, funds have discretion over what they emphasise, how they frame SI, and how much detail they provide. This makes annual reports well suited for studying disclosure choices and shifts in communication. Third, annual reports are available for virtually all funds and years in the sample because publishing an annual report is a standard reporting obligation for pension funds, which supports broad coverage and comparability over time.

At the same time, annual reports are not fully standardised. Funds differ in structure, length, language, and the use of templates or repeated text across years. These differences can introduce measurement noise and complicate direct comparisons of raw text volume or section structure. The disclosure measures constructed below are designed to reduce sensitivity to such format variation (e.g., by focusing on sentence-level SI content and by using ratio- or count-based indicators), but format heterogeneity remains a limitation of the data source.

The selection consists of all industry-wide pension funds, professional pension funds, and corporate pension funds for which annual reports are publicly available online from the overview provided by Werken aan ons Pensioen (n.d.).

To test the hypotheses, I construct text-based measures that capture multiple dimensions of SI disclosure: salience (how much SI is discussed), thematic breadth (range of SI topics covered), strategic diversity (range of SI tools/approaches discussed), portfolio integration scope (claimed coverage across assets/mandates) and concreteness/verifiability (precision of commitments).

These form the dependent variables and are inspired by Bauer, Broeders, et al. (2023), who propose disclosure measures aimed at capturing these dimensions of SI reporting. They define five measures: intensity, spectrum, variety, scope, and specificity. The remainder of this section describes how each measure is defined and constructed, and where I deviate from Bauer, Broeders, et al. (2023) to better fit the aim of this thesis (disclosure choices under contested and uncertain ESG policy signals).

Intensity (salience)

The first SI measure is Intensity, which captures how much attention a pension fund devotes to sustainable investing in its annual report. It is defined as the share of SI-related sentences among all sentences in the report. Using the SI classifier described in detail in Section 4.2, each sentence is labelled as SI-related or not. The intensity measure is then:

$$\text{Intensity}_{i,t} = \frac{N_{i,t}(\text{SI})}{N_{i,t}(\text{All})}.$$

A higher intensity value indicates that SI occupies a larger portion of the report's textual content. Because it is a share, it is less sensitive to differences in overall report length. At the same time, it captures *reported emphasis* and not the quality of SI practices.

Spectrum (thematic breadth)

Spectrum captures the breadth of SI topics covered in a given fund-year (annual report). Topic identification is performed using BERTopic, an unsupervised clustering model, as explained in Section 4.3. For this study, the topic model yields 12 SI-related topics, including reporting and regulatory frameworks (e.g., SFDR, SDGs), SI approaches (e.g., exclusions, engagement), and recurring themes (e.g., CO_2, biodiversity). Spectrum is defined as the number of distinct SI topics present in the fund-year:

$$\text{Spectrum}_{i,t} = \sum_{k=1}^{12} \mathbb{1}\{\text{topic } k \text{ is mentioned}\}.$$

Higher values indicate that SI disclosure spans more topics. This is a breadth indicator: it captures *which topics are covered*, not how extensively each topic is discussed. Because it relies on a topic model and a fixed topic set, it can be sensitive to modelling choices; I therefore interpret it as a measure of thematic coverage rather than a definitive taxonomy of SI content.

Variety (strategic diversity)

Variety measures the breadth of SI strategies that a fund reports using in a given year. Following Bauer, Broeders, et al. (2023), I consider five strategy categories: screening, ESG integration, divestment, public engagement, and private engagement. To identify which strategies appear in each annual report, I apply a rule-based dictionary approach to the set of SI-related sentences. For each strategy, I compile English and Dutch keywords commonly used by Dutch pension funds (e.g., *uitsluitingsbeleid*, *ESG-integratie*, *stembeleid*). Insights into the content of this measure are elaborated on in Section 4.4.

All SI sentences for a fund-year are concatenated and searched for these keywords. A strategy is coded as present if at least one associated keyword occurs. The variety score is therefore:

$$\text{Variety}_{i,t} = \sum_{k=1}^{5} \mathbb{1}\{\text{strategy } k \text{ is mentioned}\},$$

which indicates how many distinct SI strategy categories the fund discloses. This method is transparent and replicable, though it captures the *presence* of a strategy rather than its intensity or quality. While the dictionary used is extensive, it can miss strategies described using unusual wording and can occasionally count irrelevant mentions, as with any keyword method; I therefore treat it as a broad indicator of strategic coverage.

Scope (portfolio integration framing)

Scope measures how broadly SI commitments are framed across portfolio segments (e.g., asset classes, mandates, or geographies). In this thesis, scope is operationalised as the *number of asset classes* for which the fund explicitly describes SI commitments in its annual report. It captures the breadth of ESG framing across the portfolio rather than the portfolio's economic footprint. The asset classes tracked are equity, corporate bonds, government bonds, real estate, infrastructure, private equity, mortgages/loans, and other alternatives (fixed income securities) (De Nederlandsche Bank, 2025). A rule-based dictionary is applied to the SI-related text to code whether the fund's SI policy explicitly covers each asset class.

Let $c_{i,k,t}$ be an indicator equal to 1 if asset class k is covered by SI commitments in fund i's report in year t, and 0 otherwise. Scope is then defined as the count of covered asset classes:

$$\text{Scope}_{i,t} = \sum_{k=1}^{8} c_{i,k,t},$$

where 8 denotes the number of asset classes tracked in the analysis. Higher values indicate that SI commitments are described as applying to a broader set of portfolio segments.

The operationalisation of the Scope measure differs in an important respect from Bauer, Broeders, et al. (2023). They weight asset-class coverage by portfolio allocations to distinguish between awareness and implementation of ESG practices in a walk-versus-talk setting. In contrast, the present study focuses on strategic disclosure responses to ESG-related policy contestation and uncertainty. Accordingly, scope is measured as an unweighted count of covered asset classes, which directly captures the breadth of ESG framing across portfolio segments.

Specificity (concreteness / verifiability)

Specificity captures the extent to which SI disclosure contains concrete, verifiable commitments (e.g. targets, KPIs, time bounds). The original study by Bauer, Broeders, et al. (2023) measures specificity using paragraph-level counts. This thesis adopts a ratio-based approach using sentence-level shares rather than absolute SI-related paragraph counts. This is a deliberate methodological deviation motivated by two considerations discussed in Section 4.5.

Accordingly, the specificity metric is defined as the share of specific SI sentences among all SI-related sentences:

$$\text{Specificity}_{i,t} = \frac{N_{i,t}(\text{Spec})}{N_{i,t}(\text{SI})}.$$

A higher specificity value indicates that, conditional on discussing SI, a larger fraction of SI statements are concrete and checkable. Because it is a share, it is less sensitive to overall report length, but it does not capture whether funds talk more or less about SI in absolute terms.

Taken together, these SI measures capture distinct margins of disclosure choice. Intensity, Spectrum, Variety, and Scope primarily capture *salience and breadth* (how much and how broadly SI is discussed). Specificity captures *concreteness and verifiability* (how checkable SI statements are). These measures do not directly measure underlying SI implementation quality, but they allow an assessment of whether funds shift emphasis, topic coverage, and the concreteness of claims when ESG-related policy signals become more contested or less predictable.

3.2. Main Independent Variable: ESG Uncertainty Index

The main explanatory variable is a Dutch, media/discourse-based ESG-based Sustainability Uncertainty Index (ESGUI) developed by Ongan et al. (2025). ESGUI is available at a monthly frequency and combines ESG-related attention and uncertainty-related language using text from Economist Intelligence Unit (EIU) country reports. In plain terms, the authors construct two monthly, country-level components: (i) an ESG component based on how often ESG-related terms appear in the report and (ii) an uncertainty component based on how often uncertainty-related terms (e.g., "uncertain", "uncertainty") appear. Both components are computed as shares of total report words, to adjust for variation in report length. Each component is then rescaled to range from 0 to 100 for that country over the full sample period, and ESGUI is defined as the simple average of the rescaled ESG component and the rescaled uncertainty component (50/50 weighting).

Because ESGUI is constructed as an equal-weight composite of (i) ESG-related attention in EIU reports and (ii) uncertainty-related language, it is not a "pure" uncertainty measure. Throughout this thesis, I interpret higher ESGUI values as indicating *heightened ESG salience combined with elevated uncertainty discourse* in public discussion. The attention component captures how prominent ESG is in discussion, not whether it is discussed positively or negatively. ESGUI can therefore rise during both supportive ESG policy momentum and more contested debate or backlash, as long as ESG discussion becomes more prominent and uncertainty wording also increases. The monthly ESGUI series for the Netherlands is shown in Figure 3.1.

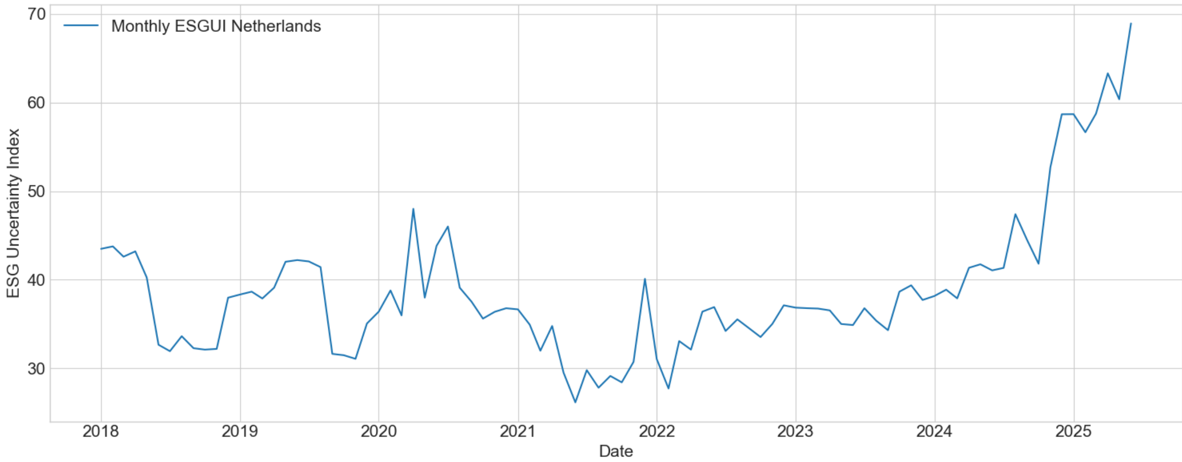

Figure 3.1: Monthly ESG Uncertainty Index (ESGUI) for the Netherlands. Notes: Index values are normalised to range from 0 (minimum) to 100 (maximum). ESGUI is an equal-weight composite of an ESG attention component and an uncertainty-language component (Ongan et al., 2025).

In this series, ESGUI increases sharply starting in 2024. One interpretation is that ESG becomes more prominent in public and supervisory discussion during periods characterised by high uncertainty in public discourse. Because ESGUI includes an attention component that is not sentiment-based, this rise should be interpreted as increased ESG salience combined with uncertainty language, rather than as a purely negative or purely regulatory shock.

A limitation of using ESGUI as a composite index is that it does not reveal which component drives the estimated relationships. A given increase in ESGUI can come from higher ESG attention, higher uncertainty language, or both. The empirical results, therefore, capture the combined effect of heightened ESG salience and uncertainty-related discourse.

3.3. Additional Variables

H2 and H3 examine whether the relationship between ESG-related contestation and uncertainty and SI disclosure depends on fund-specific characteristics. H2 focuses on *procedural visibility and organisational capacity*: more visible and better-resourced funds may be better able (or feel compelled) to maintain broader and more consistent SI disclosure when ESGUI is high. H3 focuses on *environmental sensitivity*: funds with greater exposure to carbon-intensive assets may face higher transition risk and stronger external scrutiny, potentially amplifying how ESGUI relates to their reporting choices.

To proxy procedural visibility and organisational capacity, I use fund size measured by assets under management (AUM). Larger pension funds tend to be more visible to regulators, beneficiaries, and the media, and they typically have greater internal capacity (e.g., dedicated policy, risk, and communications staff) to produce detailed disclosures. In the empirical models, AUM is treated as a time-invariant fund characteristic and enters through an interaction with ESGUI, allowing the estimated relationship to differ across funds of different size. The AUM data are sourced from De Nederlandsche Bank (DNB) and are publicly available.

To proxy environmental sensitivity, I measure each fund's portfolio carbon exposure using the weighted-average carbon intensity (WACI). WACI is expressed in tonnes of CO_2 equivalent per million euros of investee issuer revenue and is computed as the market-value (AUM)-weighted average of portfolio constituents' Scope 1 and Scope 2 emissions intensity.[1]

[1] Scope 1 and Scope 2 emissions follow the Greenhouse Gas (GHG) Protocol classification. Scope 1 refers to direct greenhouse gas emissions from sources owned or controlled by the company (e.g., onsite fuel combustion and company vehicles). Scope 2 refers to indirect emissions associated with the generation of purchased energy consumed by the company (primarily purchased electricity, heat, or steam). For an institutional investor such as a pension fund, these firm-level emissions are reflected in the portfolio through its holdings: a fund's carbon exposure is the market-value-weighted emissions intensity of the entities it invests in, such that a higher portfolio WACI indicates greater exposure to carbon-intensive investees.

Importantly, the "revenue" in the denominator refers to the investee issuer's revenue rather than the pension fund's. For exposures where issuer revenue is not well-defined (e.g., sovereign holdings), the denominator may be proxied (for example, using GDP), depending on the data provider's methodology. Overall, WACI captures the extent to which a fund's portfolio is tilted toward carbon-intensive activities and therefore serves as a proxy for baseline exposure to transition risk and carbon-related scrutiny. As with AUM, WACI is treated as time-invariant and enters the regressions through interactions with ESGUI to test whether ESGUI is more strongly associated with disclosure changes for funds with higher carbon exposure. WACI is sourced from DNB's ESG supervisory dashboard.

Because AUM and WACI are treated as time-invariant fund characteristics in this thesis, the interaction terms in H2 and H3 capture differences in the relationship across funds (e.g., large versus small, high- versus low-carbon exposure). They do not capture short-run effects of a fund becoming larger or changing its carbon intensity within the sample period.

In addition to the moderating variables, the baseline models include the pension fund funding ratio as a time-varying control. The funding ratio captures a fund's financial health and solvency position, which may influence both reporting incentives and the organisational slack available for extensive sustainability communication. For example, financially constrained funds may prioritise core reporting requirements over discretionary narrative expansion, or they may adjust the emphasis placed on SI in response to heightened stakeholder concerns. Including the funding ratio therefore helps separate the association between ESGUI and SI disclosure from concurrent changes in funds' financial conditions. Funding ratio data are sourced from DNB and are publicly available.

4 Variable Construction

4.1. Text Processing

To prepare the annual reports for text analysis, I use processing steps similar to those in Bauer, Broeders, et al. (2023). First, I convert each collected PDF into machine-readable text. The pipeline checks if each PDF has usable text or is a scanned image, and only the readable files are processed. The extraction keeps the general structure and marks simple patterns like lists or table-like lines for later recognition. The script also estimates whether the content is in English or Dutch. Metadata for the files is stored separately. This approach makes it easier to convert a large set of 785 PDFs into consistent, annotated text files and track the processing results.

After extracting text from the PDFs, I run an extra cleaning step to make the content better for analysis. This step removes PDF-related noise like headers, footers, page numbers, navigation elements, contact details, and table-of-contents fragments. It also filters out most table remnants and lines with a lot of numbers from financial overviews. Line breaks, hyphens, and spaces are adjusted so sentences and paragraphs are more continuous and readable, and common issues from annual reports and OCR are reduced as much as possible. This process turns the raw text into a more compact, readable, and consistently formatted set of documents, ready for further analysis. Some errors, like broken sentences or leftover boilerplate, may still appear, which is a known limitation of working with PDFs. Using sentence-level analysis and normalised measures, such as shares and counts, helps reduce the impact of this noise.

Next, I convert each cleaned text file into a set of well-formed sentences. The process skips any remaining navigation or table-of-contents lines and splits the text into sentences, making sure to protect abbreviations and fix common boundary errors. It removes fragments that are too short, mostly numeric, or look like headings or table rows, and joins sentences that were split by mistake.

4.2. SI Classification

To build text-based disclosure measures, I start by identifying which sentences in each annual report relate to sustainable investing (SI). This is important because SI topics appear throughout the reports along with financial, governance, and operational information. By isolating SI-related sentences, I can measure disclosure consistently, such as calculating SI intensity as the share of SI-related content, and ensure that later indicators are based on a comparable set of text.

I use a two-stage process to identify SI content, combining clear rules with a supervised language model. First, a rule-based filter flags possible SI sentences using a wide range of English and Dutch keywords and simple rules. These patterns cover common SI signals, such as ESG and sustainability terms, key topics like climate and human rights, investment-related words, and references to relevant regulations and reporting standards. Each rule-based label is saved with the rule that triggered it, making it easier to trace and check for errors. This first step aims to be broad and avoid missing SI content.

Second, I fine-tune RobBERT, a Dutch RoBERTa-based transformer[2] model, to classify sentences as SI-related versus not SI-related. BERT-style[3] models are pre-trained on large corpora and can then be adapted to specific classification tasks via fine-tuning on labelled examples. This approach is well-suited to large-scale sentence-level classification, where the objective is primarily to detect topical and contextual cues. The choice of RobBERT follows Bauer, Broeders, et al. (2023), who use the same model family for identifying sustainable-investing content in Dutch pension fund reporting.

A key methodological contribution of this thesis is how I build the training data. While Bauer, Broeders, et al. (2023) used fully manual annotation for 2,000 sentences, I test whether large language models (LLMs) can reduce the manual work without lowering label quality. In my final process, two independent LLM-assisted passes assign labels using a fixed definition and decision rules (see Appendix A.1). If both passes agree, the label is accepted; if not, I review and correct the case manually. This approach focuses human effort on the most difficult cases and produces a reliable label set for supervised fine-tuning. More details on alternative labeling strategies and robustness checks are in Appendix A.

On a stratified held-out test set of 400 sentences, the fine-tuned RobBERT classifier reached an overall accuracy of 0.978. For the SI-related class (99 sentences), precision was 0.979 and recall was 0.929, giving an F1-score of 0.953 (see Table 4.1; more details in Table A.5). This means the predicted SI sentences are very reliable, with few false positives, and most SI-related sentences are found. This provides a strong basis for calculating SI disclosure measures at the fund-year level. Any remaining classification errors would mostly add measurement noise to the dependent variables, which should weaken estimated effects rather than create false relationships.

4.3. Topic Clustering

To capture the range of sustainable investing (SI) themes discussed by pension funds, it is not enough to measure only how much they disclose or how specific their SI language is. Funds can also differ in *what* they talk about, for example climate, exclusions, engagement, or regulation. To identify these themes, a method is needed that can detect underlying topic patterns in the text without relying on predefined categories.

Topic clustering is well suited for this purpose, as it groups semantically similar sentences into coherent themes based on their language use. Unlike dictionary-based approaches, clustering allows themes to emerge in a data-driven way, while still producing interpretable categories that can be used for measurement. This makes it particularly appropriate for constructing a measure of topic breadth, which reflects the diversity of SI themes discussed in a given fund-year.

Like Bauer, Broeders, et al. (2023), this study applies BERTopic, an unsupervised topic modelling framework that combines transformer-based sentence embeddings with clustering and class-based TF-IDF[4] weighting. BERTopic first embeds sentences using a pretrained language model, then groups sentences into clusters based on semantic similarity, and finally extracts representative keywords for each cluster. Importantly, BERTopic operates without labelled training data and does not rely on predefined topic taxonomies, in contrast to the supervised RoBERTa models used earlier in the analysis to classify SI-related sentences and their specificity. While the RoBERTa models are trained to recognise known categories, BERTopic is designed to discover structure in the data itself.

The topic model is trained exclusively on sentences previously classified as SI-related, in this study a total of around 64,000 sentences, ensuring that all resulting clusters pertain to sustainable investing rather than general financial reporting. The initial BERTopic run produces a larger set of clusters, including substantive SI themes as well as residual clusters reflecting generic boilerplate, regulatory templates,

[2]A transformer model is a type of neural language model that represents text by learning relationships between all words in a sentence simultaneously, rather than processing them sequentially (Vaswani et al., 2017).

[3]BERT (Bidirectional Encoder Representations from Transformers) is a neural language model that learns contextual word representations by processing text bidirectionally, meaning it considers both left and right context simultaneously. It is pre-trained on large unlabeled corpora using tasks such as masked language modelling and next sentence prediction, and can then be fine-tuned for downstream classification tasks with relatively limited labelled data (Devlin et al., 2019).

[4]TF-IDF (term frequency-inverse document frequency) is a weighting scheme that assigns higher weight to words that occur frequently within a given document or cluster but relatively rarely across the wider corpus, thereby highlighting terms that are characteristic of that document/cluster. In BERTopic, a class-based variant (c-TF-IDF) is used, where all texts assigned to a topic are treated as one "class" to extract topic-representative keywords.

Figure 4.1: Word cloud of final selection of relevant topics with sizes relative to their prevalence in all annual reports.

or linguistic artefacts. Following inspection of the top keywords and representative sentences per cluster, non-substantive clusters are excluded, and closely related clusters are merged where appropriate. This results in a final set of twelve distinct, substantively meaningful SI topics, covering environmental, social, governance, regulatory, and stewardship dimensions of sustainable investing. The selection of topics is shown in Figure 4.1.

Because topic models can be sensitive to modelling choices (e.g., embedding model settings, clustering parameters, and the merging of clusters), I interpret the resulting topics as an empirical representation of *thematic coverage* in this corpus rather than as a definitive taxonomy. The main goal is to measure whether funds cover a broader or narrower set of SI themes over time, using a fixed and transparently documented procedure.

CO_2 is the most prevalent term in the topic-model training data, appearing more than 2,500 times. *Biodiversity* is the least prevalent topic, but still occurs over 450 times. Figure 4.2 shows the number of pension funds covering selected topics in their annual report over time. These topics are a subset of the total 12 topics. A noticeable pattern is that most topics become more commonly covered over time. *SFDR* increases steeply after 2019, consistent with the introduction and roll-out of the framework.

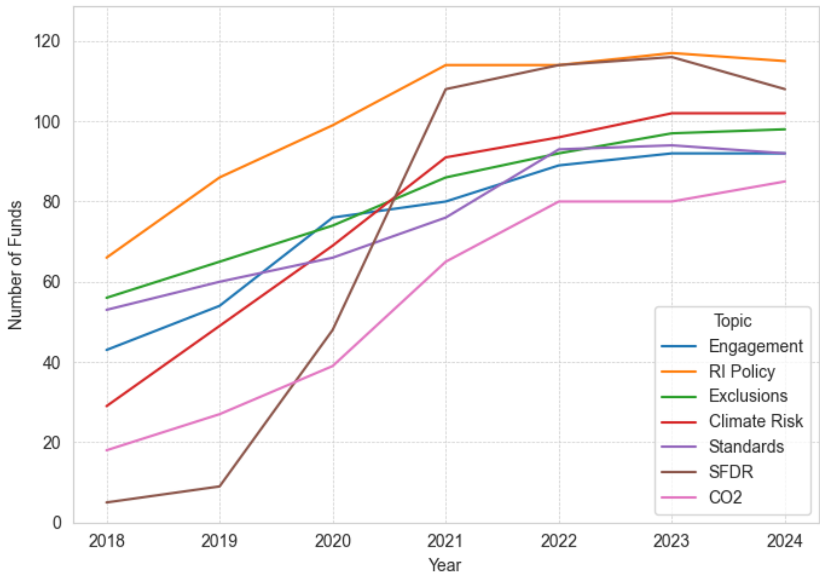

Figure 4.2: Number of pension funds discussing selected sustainable investing topics over time.

4.4. Variety and Scope Construction

The construction of the Variety and Scope measures is more straightforward and uses a rule-based approach. Variety counts which SI strategies are mentioned in a fund's SI-related sentences, and Scope counts for which asset classes the fund frames SI as applying. There are 5 SI strategies in total and 8 asset classes. Figures 4.3 and 4.4 provide more information on the content underlying the Variety and Scope measures.

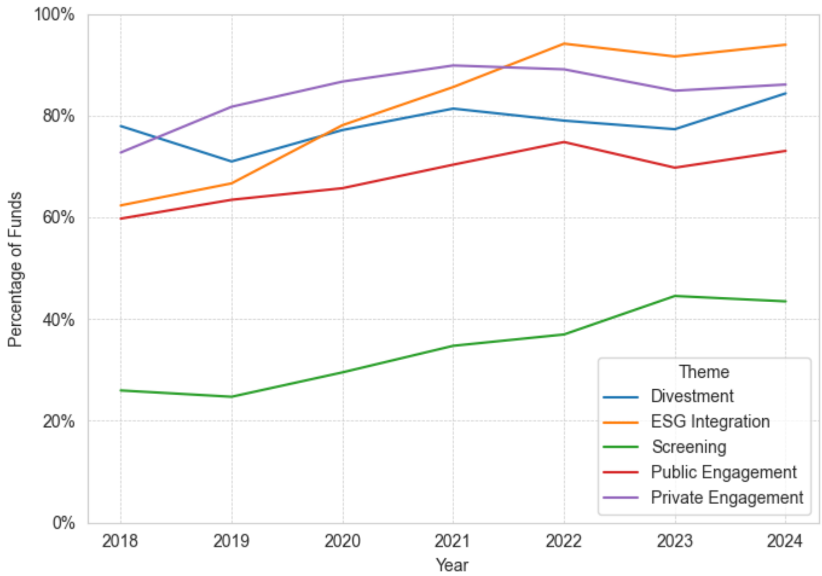

Figure 4.3: Variety measure over time. The percentages represent the fraction of pension funds that mention a certain sustainable investing (SI) strategy in their annual reports over time.

Figure 4.3 shows the fraction of pension funds that mention a certain SI strategy over time. Whereas Bauer, Broeders, et al. (2023) find divestment to be the most popular strategy, my results show an increase in ESG integration, with nearly all pension funds from the sample covering this strategy in their annual reports. Both types of engagement show a gradual increase over time. Overall, all strategy categories become more commonly mentioned.

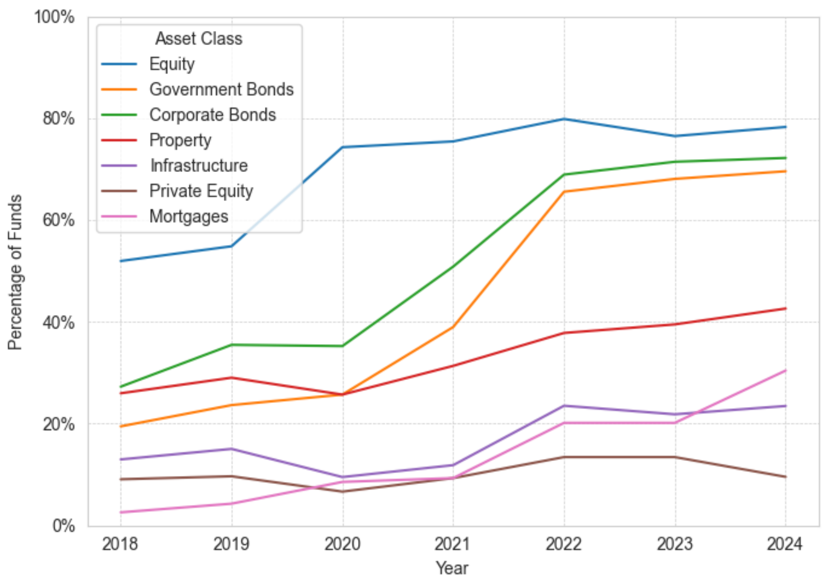

Figure 4.4: Scope measure over time. The percentages represent the fraction of pension funds that frame their sustainable investing (SI) policy as covering a certain asset category over time.

Likewise, Figure 4.4 shows the share of analysed funds that cover a specific asset class in their SI policy statements. Most asset classes show an increase over the years. Like Bauer, Broeders, et al. (2023), equity remains the most common asset class for SI framing, which is plausible because several SI approaches (e.g., exclusions, screening, and engagement) are relatively straightforward to describe for listed equity. Corporate and government bonds also become more commonly covered after 2020, which may reflect both changing practices and improved data availability and reporting conventions in these categories.

4.5. Specificity Classification

The next classification task concerns the *specificity* of SI-related sentences. Identifying SI-related content is a necessary first step, but it does not capture whether a statement provides substantive, decision-relevant information or remains vague and generic. In sustainability reporting, funds may mention SI-related topics in broad terms (e.g., commitments or general principles) without disclosing concrete actions, targets, timelines, or measurable outcomes. Classifying specificity therefore helps distinguish more informative disclosure from boilerplate language and enables a richer assessment of disclosure beyond mere SI volume.

This dimension is also used by Bauer, Broeders, et al. (2023). In this thesis, specificity is determined at the *sentence* level rather than the paragraph level, and I operationalise it as a *specificity ratio*: the fraction of SI sentences that are classified as specific. Sentence-level units are less affected by PDF extraction artefacts than paragraph boundaries and match the short-text strengths of transformer models. For the econometric analysis, a ratio also captures changes in the *composition* of SI disclosure (more versus less checkable statements) rather than changes in overall report length.

In contrast to SI detection, specificity is not primarily a topical task. It concerns the concreteness and verifiability of claims and therefore requires semantic judgement about whether a sentence describes an attributable action, measurable outcome, explicit target, or procedural obligation. As a result, prompt-based batch annotation in Agent mode (which was used for the first classification task) proved unstable for this task, with substantial variation across repeated runs. I therefore use independent, per-sentence classification via API calls, which yields substantially more reproducible outputs under a fixed prompt and deterministic decoding.

A further practical challenge is class imbalance: within the SI corpus, specific statements are relatively rare. To obtain a meaningful evaluation of minority-class performance, I construct a hand-labelled gold set that combines random sampling with targeted oversampling of likely specific sentences. This gold set is then used to evaluate the API-based labelling procedure and to guide the construction of a balanced training set for the final specificity classifier (full details in Appendix A.2).

Table 4.1: Performance of sentence-level classifiers

	SI sentence classifier	**Specificity classifier**
Task	SI-related vs. not SI-related	Specific vs. not specific
Unit of classification	All sentences	SI-related sentences
Test set size (N)	400	250
Positive-class support	99 (SI-related)	93 (Specific)
Accuracy	0.978	0.896
Precision (positive class)	0.979	0.876
Recall (positive class)	0.929	0.839
F1-score (positive class)	0.953	0.857

Notes: “Positive class” refers to SI-related sentences in the SI classifier and to specific SI sentences in the specificity classifier. Both models are evaluated on held-out test sets; full classification reports and additional diagnostics are provided in the appendix.

The final specificity classifier achieves strong performance on a held-out test set (Table 4.1). Overall accuracy is 89.6%. For the “specific” class, precision is 0.876 and recall is 0.839 (F1 = 0.857), indicating that most predicted specific statements are correct while a limited share of truly specific sentences is missed. These results support the use of the resulting specificity ratio as a scalable operational measure of disclosure precision at the fund-year level.

4.6. ESGUI Exposure Construction

This section describes how the monthly ESG Uncertainty Index (ESGUI) is mapped to the annual report panel. Let i index pension funds and t index reporting years. Each fund-year observation is linked to the publication month of the corresponding annual report, denoted $p(i,t)$. Figure 4.5 shows that publication is clustered: most funds publish the annual report for year Y in late spring/early summer of $Y+1$, with a strong peak in June.

The key design choice is to align uncertainty exposure with the information environment under which the report text is produced. Annual reports are drafted, revised, and approved over multiple months. Using a single-month ESGUI value at publication would therefore overstate high-frequency noise, while a single annual national ESGUI value would be too coarse and would not reflect fund-specific publication timing. I therefore construct ESGUI exposure as a publication-aligned rolling mean of the monthly national ESGUI series. The baseline measure is the average ESGUI over the 12 months ending in the publication month:

$$ESGUI_{it}^{(12)} = \frac{1}{12}\sum_{m=0}^{11} ESGUI_{p(i,t)-m}. \tag{4.1}$$

This rolling-window approach is preferred for two related reasons. First, it offers a closer conceptual match to the reporting process by capturing a persistent ESG discourse environment rather than a point-in-time reading. Second, it produces transparent within-year variation: because each fund's exposure window ends in its own publication month, funds publishing earlier versus later in the reporting season are mapped to slightly different ESGUI histories even though the underlying ESGUI series is common to all funds. A note however, is that within each year, dispersion across funds is comparatively modest because all funds share the same national ESGUI series and publication months are concentrated in a narrow reporting season (Figure 4.5).

I use a 12-month window as the baseline because it matches the annual reporting cycle and reflects ongoing conditions that likely influence disclosure focus and scope. Shorter windows, like 6 or 9 months, tend to highlight short-term changes close to the publication date and can be more affected by timing noise. Another option would be to leave out a brief period just before publication, assuming the report text is finalised months in advance. However, the last few months before publication usually include compliance checks, risk reviews, board approval, and final edits, when disclosure language and framing can still change. That is why our baseline window covers the months right before publication. We report timing robustness checks using different rolling windows and cut-off points in the Results section and Appendix A.3.

These factors are important for interpreting the fixed-effects design. The main regressions use both fund and year fixed effects. Fund fixed effects control for differences in reporting style and disclosure habits that do not change over time, while year fixed effects control for shocks and policy changes that affect all funds in a given year. In this setup, the coefficient on ESGUI is mainly identified by differences in exposure within the same year, which are caused by publication-month differences due to the rolling-window method. Because of this, it is necessary to show the distribution of publication months and check that the results are not just a result of the rolling-window approach. The robustness tests discussed later, including changes in timing and placebo publication-month assignments, address this issue directly.

4.7. Sample Overview

Before discussing the empirical design and results, Table 4.2 gives a final overview of the sample, summarising descriptive statistics for all variables used in the next chapter's models. The analysis covers 785 annual reports from 124 pension funds between 2018 and 2024.[5] After removing observa-

[5] 83 reports were missing from the selection made. The panel is therefore not completely balanced. However, only funds are included that had more than one report from consecutive years available.

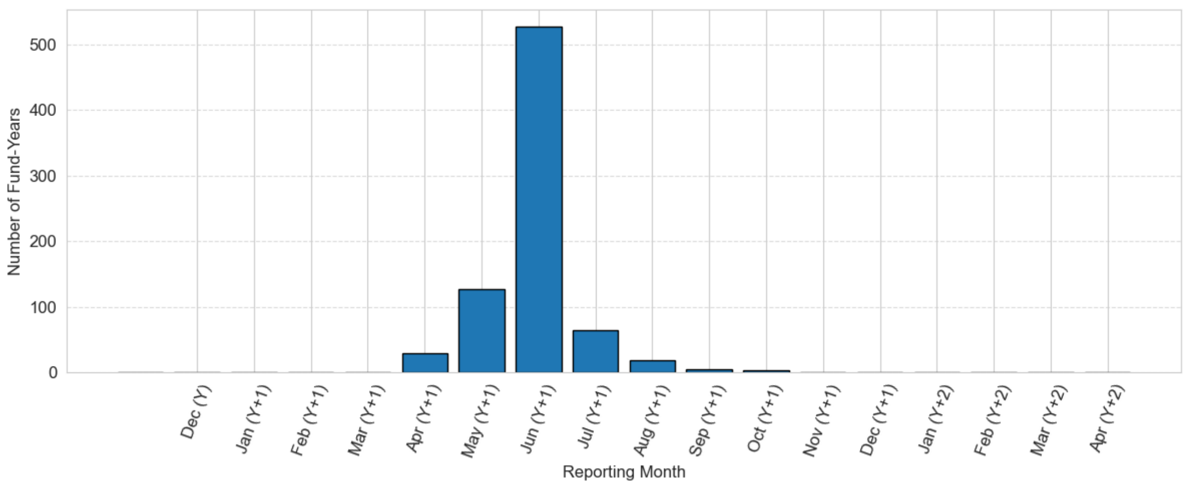

Figure 4.5: Distribution of annual report publication months relative to the reporting year. Each observation represents a fund-year. *Note:* The notation `(Y)` indicates the reporting year, while `(Y+1)` and `(Y+2)` indicate publication in the first and second calendar year after the reporting year, respectively.

tions with missing values in any baseline variables, the main regression sample includes 781 fund-year observations, or 727 when WACI is included.

There are several important points in Table 4.2. First, report length varies a lot between fund-years. The median report has 1,533 sentences, and the top quarter have more than 2,190 sentences. Because of this, raw counts of SI sentences and breadth measures (like Spectrum, Variety, and Scope) may partly reflect how long or detailed the reports are. To address this, the analysis uses normalised outcomes (Intensity and Specificity as ratios) and, when needed, includes controls or checks for report length in the regression models.

Second, the amount of SI-related content varies a lot between funds. Most reports have about 3% SI-classified sentences, or around 45 sentences, but the range is wide. Some reports have almost no SI content, while others have nearly 30% SI sentences. This variation also appears in Specificity, which is a ratio that shows a lot of spread and is skewed to the right. This skewness means that a small number of highly detailed fund-years could affect the results, so using fixed effects and robustness checks in the analysis is important.

Third, the breadth measures show meaningful differences in what funds communicate about sustainable investing. On average, a fund-year mentions about 6 to 7 out of 12 possible SI topics (Spectrum), around 3 to 4 out of 5 strategy categories (Variety), and about 3 out of 8 asset classes (Scope), with considerable variation across the sample. This pattern suggests that funds differ not only in the amount of SI discussion but also in its thematic and strategic breadth, supporting the need to analyse multiple outcome dimensions instead of relying on a single disclosure measure.

Finally, the explanatory and control variables also show a lot of variation. ESGUI changes meaningfully over the sample period, while AUM is highly skewed with a few very large funds. The funding ratio is more concentrated but still varies across fund-years. These features explain why the empirical design focuses on within-fund variation over time and controls for time-varying fund characteristics, while being careful with cross-sectional differences.

Table 4.2: Descriptive statistics of variables used in the empirical analysis

Variable	Description	Unit	Obs	Mean	Std. Dev.	Min	P25	P50	P75	Max	Source
Panel A: Dependent variables (SI disclosure outcomes)											
Intensity	SI sentences as share of total sentences	Ratio (0-1)	781	0.04	0.03	0.00	0.02	0.03	0.05	0.29	Annual reports
Spectrum	Number of SI topics mentioned	Count (0-12)	781	6.68	3.20	0.00	4.00	7.00	10.00	12.00	Annual reports
Specificity	Specific SI sentences as share of SI sentences	Ratio (0-1)	781	0.09	0.07	0.00	0.04	0.09	0.13	0.43	Annual reports
Variety	Number of SI strategies mentioned	Count (0-5)	781	3.49	1.32	0.00	3.00	4.00	4.00	5.00	Annual reports
Scope	Number of asset classes with SI discussion	Count (0-8)	781	3.15	2.31	0.00	1.00	3.00	5.00	8.00	Annual reports
Panel B: Main independent variable											
ESGUI (12m mean)	ESG uncertainty (12-month rolling mean, publication-aligned)	Index (0-100)	781	38.88	7.30	30.48	34.50	36.95	38.59	57.75	Ongan et al. (2025)
Panel C: Controls and report characteristics											
Funding ratio	Funding ratio (divided by 10; one unit = 10 percentage points)	10 pp	781	11.67	1.54	7.90	10.61	11.54	12.58	21.86	DNB
Total sentences	Total number of sentences in the annual report	Count	781	1,894	543	639	1,533	1,807	2,190	4,994	Annual reports
Panel D: Moderators (heterogeneity analyses)											
AUM	Assets under management	EUR bn	781	13.61	56.11	0.02	0.66	1.80	5.77	531.99	DNB
WACI	WACI (Scope 1+2)	(tCO_2e / €m issuer revenue)	727	72.4	26.5	1.23	53.0	72.4	85.9	220.8	DNB

Notes: Descriptive statistics are computed for all non-missing observations for each variable. WACI is available for a smaller subsample (727 fund-years). In the heterogeneity specifications, fund size and carbon exposure enter as $\log(AUM)$ and $\log(WACI)$ (mean-centered for interpretability). ESGUI is sourced from Ongan et al. (2025). DNB refers to De Nederlandsche Bank.

To give a clearer view of how SI measures change over time, Figure 4.6 shows their distributions by year. Each panel displays one indicator for a specific year, with a red dotted line marking the median. Across all five measures, the median rises over time, which means that SI disclosure generally increases each year.

The shapes of the distributions differ between indicators. Intensity and Specificity are right-skewed each year, with most fund-years at low values and a few with much higher scores. Variety and Scope have wider and more varied distributions, showing differences in how many SI themes are included and how broadly SI policies cover asset classes. Both measures shift toward higher values over time.

The figure shows that SI disclosure varies a lot from year to year, but overall, it has steadily increased during the sample period. Since 2018, funds have reported more often, with more detail and on a wider range of SI topics, strategies, and asset classes. This increase probably comes from sector-wide changes, like more regulatory focus on sustainability and new reporting practices. To make sure these general trends are not mistaken for ESG uncertainty, the analysis uses year fixed effects to control for changes that affect all funds in a given year.

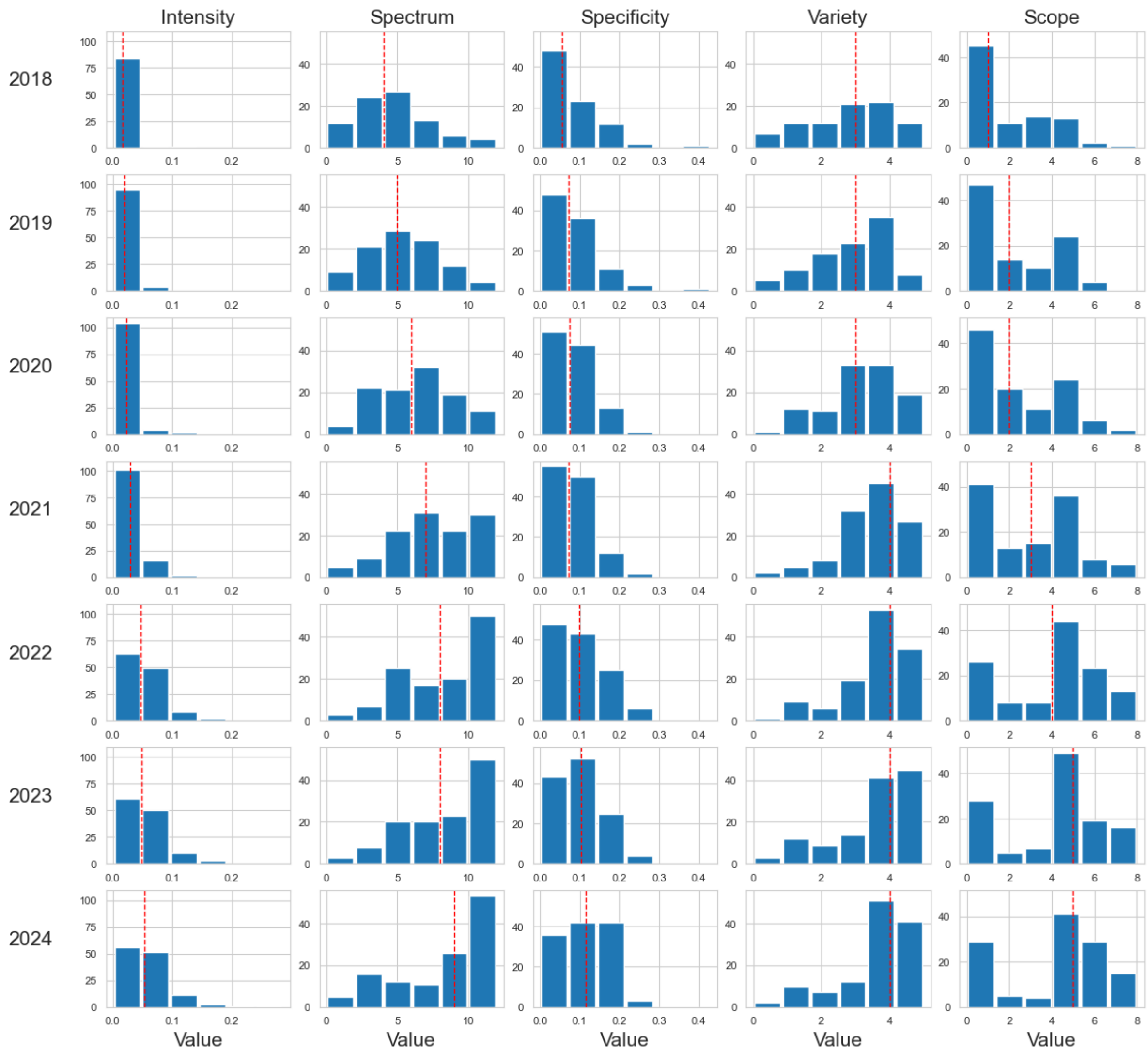

Figure 4.6: Distribution of SI measures. The figure shows the distribution of the different SI measures for all analysed years. Every plot visualises the distribution of a particular measure in a specific year. The dotted red line in each plot represents the median.

Figure 4.7 presents basic Pearson correlations between the main variables. The disclosure outcomes are all positively related, especially the breadth measures (spectrum, variety, and scope), suggesting they reflect a shared dimension of more extensive disclosure. Report length is closely linked to these breadth measures, as longer reports allow for more topics, strategies, or asset classes to be mentioned. Fund size is also strongly linked to report length and is positively related to the disclosure measures, which makes sense since larger funds can report more.

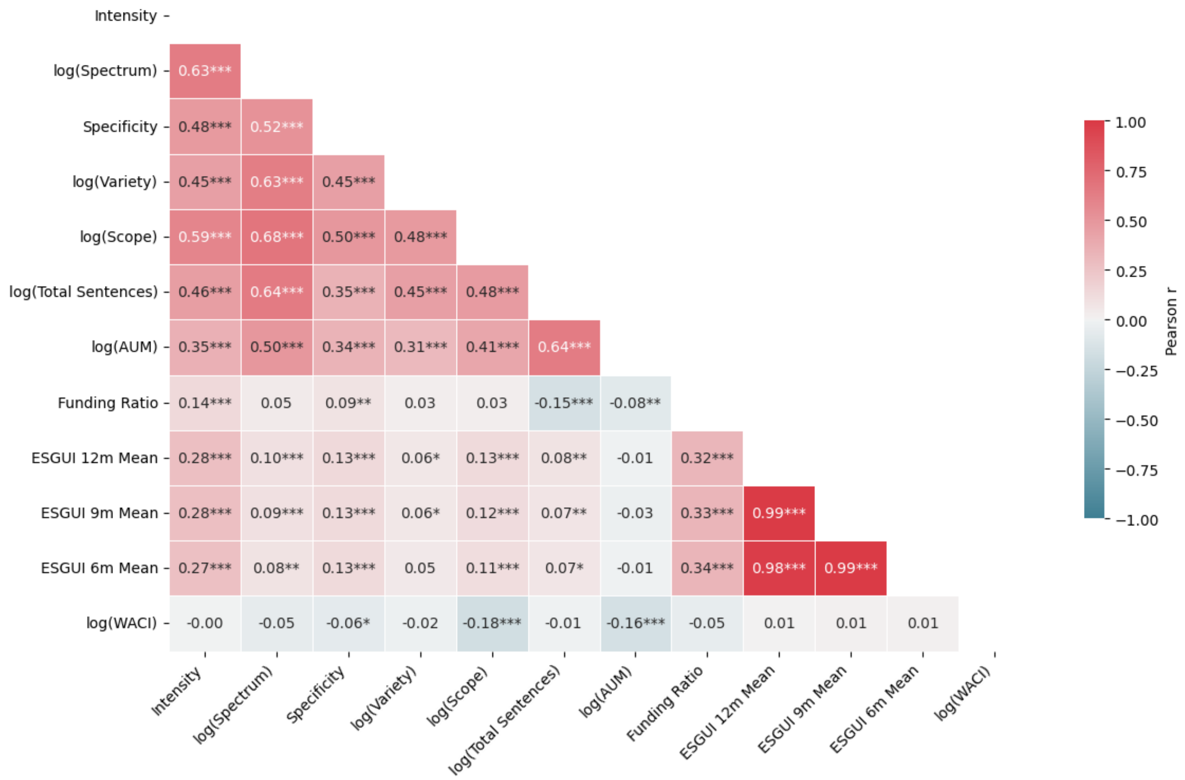

Figure 4.7: Correlation Heatmap. Pearson correlation coefficients (r) showing the strength and direction of linear relationships between study variables. Significance levels: *p < 0.1, **p < 0.05, ***p < 0.01.

In the main regressions, I do not include controls that might be part of the causal link between uncertainty and disclosure or that could change what the outcome measures. Report length is an important example. If uncertainty influences report length, or if other factors affect both length and disclosure, then controlling for length changes the focus of the model. Because of this, models that adjust for report length are used as sensitivity checks, not as the main approach. This is also considered in the robustness tests.

5

Empirical Design and Results

I study whether periods in which ESG-related issues are more salient in uncertain environments are associated with changes in how Dutch pension funds discuss sustainable investing (SI) in their annual reports. The main challenge is that funds are quite distinct. Some funds are structurally more "sustainability-heavy" in their reporting: they have more resources, more reporting capacity, and often longer reports. To avoid confusing these permanent differences with the role of uncertainty, I use fund fixed effects. This means I compare each fund to itself over time, after removing anything that is stable for that fund across the sample.

A second challenge is that the main explanatory variable is national. The ESG-relevant environment is proxied using a Dutch discourse-based index (ESGUI), which reflects periods in which ESG-related issues are discussed alongside elevated uncertainty in public discourse. In a given month, all funds face the same national ESGUI level. There is therefore no obvious treated group and control group. Instead, the key idea in this thesis is that exposure can differ once I take *timing* into account.

I construct a fund-specific "writing window". For each fund-year, I take the twelve months before the fund publishes its annual report and average ESGUI over that period. This gives the publication-aligned exposure measure $ESGUI_{it}^{(12)}$. The intuition is as follows: the annual report is drafted, discussed, revised, and finalised before publication. The uncertainty environment in the months leading up to publication is therefore the environment in which reporting decisions are made. Because funds publish in different months, their twelve-month windows cover different sets of calendar months. This leads to different average exposure levels even within the same report-year. In this setup, differences in rolling-window averages play the role of a continuous "treatment intensity".

A second advantage of this design is that I can include year fixed effects. These absorb broad developments that affect all funds within a report-year, such as common changes in regulation, market conditions, or overall trends in ESG attention. This is important because ESGUI itself is a national series: if I only used the national level of ESGUI without publication alignment, year fixed effects would soak up the relevant variation. With publication alignment, however, exposure still differs across funds within the same report-year, so year fixed effects do not eliminate the identifying variation.

With fund and year fixed effects, identification therefore comes mainly from *within-report-year* differences in exposure induced by publication timing. In practice, this means the design mainly compares funds observed in the same report-year whose publication dates place them at different points on the monthly ESGUI path, which implies slightly different rolling-window averages.

Two details about where variation comes from are useful for interpretation:

- **Cross-fund timing differences within a year.** Even if each fund publishes in a stable month every year, early versus late publishers in the same report-year will generally have different $ESGUI_{it}^{(12)}$ because their writing windows begin and end in different months.

- **Within-fund shifts in publication month.** If a fund publishes earlier or later than it usually does in a given year, its writing window shifts and the average ESGUI over that window can change.

In the data, publication timing is fairly stable: 73.7% of funds publish in the same calendar month as in the previous year. This suggests that most identifying variation comes from cross-fund timing differences within the same report-year, rather than from frequent month switching within funds.

Hypothesis 1 is tested with an unbalanced fund-year panel and a fixed-effects model that relates each disclosure outcome $Y_{it}^{(k)}$ to the publication-aligned ESG Uncertainty Index $ESGUI_{it}^{(12)}$, controlling for the funding ratio and including fund and year fixed effects:

$$Y_{it}^{(k)} = \beta\, ESGUI_{it}^{(12)} + \delta\, FR_{it} + \mu_i + \gamma_t + \varepsilon_{it}, \tag{5.1}$$

where $k \in \{$*intensity*, *spectrum*, *specificity*, *variety*, *scope*$\}$, μ_i are fund fixed effects, and γ_t are year fixed effects. FR_{it} is the funding ratio of fund i at year t. The coefficient β measures how SI disclosure moves with publication-aligned exposure after removing (i) stable differences between funds and (ii) shocks that affect all funds in the same report-year. Because $ESGUI_{it}^{(12)}$ is based on the twelve months *before publication*, funds in the same report-year can still have different exposure levels if they publish in different months. Thus, with fund and year fixed effects, β is identified primarily from within-report-year exposure differences created by publication timing.

Unless stated otherwise, I use standard errors that are robust to heteroskedasticity[6] and that are robust to potential correlation of the error term along two dimensions: within funds over time and across funds that publish in the same period. Errors may be correlated within funds because reporting routines and resources are persistent. They may also be correlated across funds that publish in the same month because publication is seasonal and ESGUI is constructed at a monthly frequency. To account for both possibilities, the baseline results use *two-way clustered* standard errors by (i) fund and (ii) publication year-month.

Governance variables are not included as baseline controls. Although Bauer, Broeders, et al. (2023) show that governance characteristics (e.g. board size, tenure, gender composition) can matter for SI disclosure, many of these differences are broadly stable and are already captured by fund fixed effects. In addition, governance can change slowly over time and may respond to broader ESG pressures, which means it could be affected by the same environment that drives uncertainty and disclosure. For these reasons, I do not include governance variables in the main model; they are better interpreted as potential channels for future research.

This publication-aligned design assumes that exposure to the ESG-relevant environment during the drafting period comes before any disclosure choices in the annual report. If we interpret this causally, future exposure should not influence content that has already been written. Another issue is the possible endogeneity of publication timing. I address both of these assumptions before presenting the main results.

To check the timing, I run a placebo test where the exposure variable is based on the average ESGUI from the twelve months after the publication month. If future exposure predicts current disclosure outcomes, it would mean that the baseline estimates might be picking up ongoing trends or missing factors, rather than exposure during the writing period.

The placebo specification is estimated on a restricted sample because the ESGUI series ends in May 2025. This cutoff means we cannot create a full 12-month post-publication window for reports published near the end of the sample ($N = 535$). Since publication timing is usually consistent within funds, this restriction tends to exclude more of the later publishers. As a result, the placebo and baseline coefficients should not be directly compared, since some of the differences reflect changes in the sample rather than only differences in the estimated relationships.

Table 5.1 compares the baseline specification using pre-publication exposure (Panel A) to a placebo exposure measure based on the 12 months *after* publication (Panel B), estimated on the same restricted

[6] Standard errors that are robust to heteroskedasticity do not assume the variance of the regression error term is the same for all observations. This means the estimated uncertainty around the coefficients stays valid even if the spread of the unexplained part of the outcome variable changes across funds or over time. This situation often happens in panel data when reporting practices differ.

sample. For *Intensity*, *Variety*, *Scope*, and *Specificity*, the placebo coefficients are small and statistically insignificant, which is consistent with the writing-window interpretation: future ESG uncertainty does not systematically predict current disclosure content.

Table 5.1: Baseline vs Placebo: Pre- vs Post-Publication ESG Uncertainty (Restricted Sample)

	(1)	(2)	(3)	(4)	(5)
Dep. var.	Intensity	Spectrum	Variety	Scope	Specificity
Panel A: Baseline (pre-publication ESGUI)					
ESGUI (12m pre)	-0.001	-0.185	-0.117**	-0.107	-0.007**
	(0.001)	(0.121)	(0.047)	(0.108)	(0.003)
Funding Ratio	-0.002	0.050	0.258**	-0.142	-0.000
	(0.002)	(0.231)	(0.121)	(0.184)	(0.009)
Panel B: Placebo (post-publication ESGUI)					
ESGUI (12m future)	-0.001	-0.339***	-0.034	-0.020	0.002
	(0.001)	(0.063)	(0.047)	(0.088)	(0.004)
Funding Ratio	-0.002	0.052	0.263**	-0.138	-0.000
	(0.002)	(0.236)	(0.125)	(0.187)	(0.009)
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	535	535	535	535	535

Notes: Panel A reports the baseline specification using publication-aligned ESG uncertainty measured over the 12 months before publication. Panel B reports a placebo using the mean ESGUI over the 12 months after publication. Both panels are estimated on the same restricted sample; the sample is smaller because the ESGUI series ends in May 2025. All specifications include fund and year fixed effects and control for the funding ratio. Standard errors are two-way clustered by fund and publication year-month. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Spectrum is the main exception. The placebo coefficient for Spectrum is negative and statistically significant, but the pre-publication estimate in the same restricted sample is not. This pattern suggests that the Spectrum relationship may partly reflect slow-moving ESG discourse cycles or ongoing reporting habits that move together with ESGUI before and after publication. Because of this, I interpret the Spectrum results with more caution and focus more on outcomes where the placebo test gives stronger reassurance.

Another concern is that publication timing might be influenced by other factors. For example, if funds systematically adjust publication dates in response to ESG uncertainty or similar considerations, then publication-aligned exposure could partly reflect these timing decisions rather than the reporting environment during drafting. Appendix Table A.10 examines whether ESG uncertainty is associated with changes in publication timing. The results mostly suggest that there are no systematic shifts in how late reports are published, although there is weak evidence that higher ESGUI is linked to a slightly higher probability of changing the publication month. Because publication timing and the measurement of uncertainty exposure could affect identification, the analysis also includes robustness checks discussed below and reported in the Appendix.

5.1. The Effect of ESG Uncertainty on SI-Disclosure

The theory in Chapter 2 suggests two main ways funds adjust when ESG-related issues become more important in uncertain environments. First, funds might talk less about sustainable investing (SI) or cover fewer SI topics, strategies, or asset classes (contraction; H1a). Second, if they continue to discuss SI, they may use less concrete and less verifiable language (within-SI hedging; H1b). This section examines these two adjustments using the disclosure measures from Chapter 3.

Since the baseline model includes year fixed effects, the ESGUI coefficient reflects differences in exposure within the same report year, based on when reports are published. These differences are fairly small (see Appendix Table A.11), so the results should be seen as local responses to small changes in ESGUI exposure within a year. Larger changes across years are absorbed by the year fixed effects and are not directly measured in this setup.

The main results in Table 5.2 show a contraction pattern. When ESGUI exposure is higher at the time of publication, funds show lower Intensity and cover a narrower range in Spectrum, Variety, and Scope. In contrast, the effect on Specificity is small and not statistically significant in the full sample. This suggests that greater ESG-related uncertainty mainly leads to less prominent and less broad SI disclosure, rather than clear changes in how concrete the language is.

Table 5.2: Baseline Results: ESGUI Exposure and SI Disclosure Outcomes

	Salience and Breadth				Concreteness / Verifiability
	Intensity (1)	Spectrum (2)	Variety (3)	Scope (4)	Specificity (5)
ESGUI (12m)	-0.002*	-0.145**	-0.069*	-0.137**	-0.003
	(0.001)	(0.071)	(0.036)	(0.062)	(0.002)
Funding Ratio	-0.005*	-0.088	0.085	-0.156	-0.006
	(0.003)	(0.160)	(0.122)	(0.165)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
Within R^2	0.022	0.008	0.006	0.011	0.007

Notes: The table reports coefficients from the baseline fixed-effects specification in Equation 5.1. Standard errors (in parentheses) are two-way clustered by fund and publication year-month. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-fund variation explained by the included regressors. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Since the baseline coefficients come from small differences in exposure within each report year, it helps to express them in more practical economic terms. In this sample, the standard deviation of residualised $ESGUI^{(12)}$ is about 0.85 ESGUI points (Appendix Table A.11). Adjusting the baseline estimates to match this within-year variation shows that the effects are modest in absolute terms.

For *Intensity*, a one-standard-deviation increase in ESGUI exposure within a year is linked to a drop of about $-0.002 \times 0.85 \approx -0.0017$ in the SI share of the report. Compared to the median intensity of 0.03 (Table 4.2), this means a reduction of about 5.7%. Since the median report has around 1,800 sentences, this works out to about three fewer SI-related sentences in a typical report.

For the breadth measures, the effects within each year are also small. A one-standard-deviation increase in ESGUI exposure leads to about -0.12 fewer SI topics (*Spectrum*), -0.06 fewer SI strategy categories (*Variety*), and -0.12 fewer asset classes with SI discussion (*Scope*). Compared to the sample means in Table 4.2, these changes are reductions of about 1.8%, 1.7%, and 3.7%, respectively.

For *Specificity*, a one-standard-deviation increase in ESGUI exposure results in an estimated $-0.003 \times 0.85 \approx -0.0026$ decrease in the share of SI sentences with concrete, verifiable elements. Compared to the median specificity level of 0.09, this is a reduction of about 2.9%. With an average of 82 SI sentences per report, this means about 0.2 fewer specific SI sentences. However, since the estimated coefficient is not statistically significant, this effect should be viewed with caution.

These calculations show that the estimated effects do not represent major changes in reporting strategy. Instead, they suggest small adjustments in disclosure, like leaving out a topic, mentioning fewer strategies, or discussing one less asset-class application of SI when ESG becomes more prominent in uncertain public discussions. Importantly, these results likely reflect changes in *disclosure behaviour* rather than actual SI implementation.

Hypothesis 1a asks if funds make SI disclosures less prominent and cover fewer topics when ESG issues become more important in uncertain times. I test this using Intensity (SI salience), Spectrum (topic breadth), Variety (strategy breadth), and Scope (asset-class breadth).

The main results show that when ESGUI exposure is higher, funds spend a bit less space on SI and cover fewer SI topics, strategies, and asset classes. This suggests that funds mostly narrow their focus and coverage, rather than stop SI reporting entirely.

Looking at the different measures, Scope stays fairly stable, while Variety is more sensitive. However, the earlier placebo test shows that future ESGUI exposure also links to Spectrum, which makes the timing of this result less clear. This may mean that the Spectrum finding partly reflects ongoing ESG trends or reporting habits. Overall, higher ESGUI exposure is linked to a modest reduction in SI salience and breadth, with the strongest evidence for Intensity, Scope, and to a lesser extent, Variety. The Spectrum result supports this but is not conclusive.

Hypothesis 1b examines whether, conditional on still discussing SI, funds shift toward less concrete and less verifiable language when ESGUI exposure is higher. This margin is measured by *Specificity*, defined as the share of SI sentences that contain concrete, verifiable elements such as actions, targets, metrics, or timelines.

The baseline results provide limited evidence for such a shift. The estimated coefficient on *Specificity* is small ($\beta = -0.003$) and statistically indistinguishable from zero. Economically, a one-standard-deviation increase in ESGUI exposure would correspond to a decline of approximately $-0.003 \times 0.85 \approx -0.0026$ in the share of SI sentences containing concrete elements, which is modest in magnitude.

Overall, the results do not support H1b. The main change with higher ESGUI exposure is in whether and how broadly SI is discussed, not in how concrete the remaining SI language is.

In summary, the evidence supports H1a (contraction margin) more than H1b (within-SI hedging). When ESG issues become more important in uncertain times, Dutch pension funds talk about SI less and cover it less broadly. There is little sign that the SI content that remains becomes less concrete or verifiable.

Substantively, the results show that companies tend to leave out certain information or limit what they disclose, instead of making their language about SI topics more vague. However, these effects are not very large. For example, a one-standard-deviation increase in ESGUI exposure leads to only a few fewer SI-related sentences and slightly fewer disclosure categories in a typical report.

These findings are best seen as small but consistent changes in how much and how broadly funds disclose SI, not as major shifts in sustainable investment practices or communication. This matters for supervisors: annual reports may adjust slightly to uncertain ESG environments even if actual investment practices stay the same.

5.2. Moderation by Fund Size (H2)

Hypothesis 2 examines whether the effect of ESG-related contestation/uncertainty on SI disclosure depends on fund visibility and capacity, proxied by fund size. Because AUM is observed as a time-invariant fund characteristic in the available data, the models in this section do not identify how a fund changes as it grows. Instead, they test whether *different types of funds* (larger versus smaller) have different disclosure *slopes* with respect to $ESGUI_{it}^{(12)}$. Since AUM is time-invariant, its main effect is absorbed by fund fixed effects, and H2 is tested through an interaction with $ESGUI_{it}^{(12)}$.

Let AUM_i denote fund size and define $logAUM_i = \log(AUM_i)$. To facilitate interpretation, $logAUM_i$ is mean-centred:

$$logAUM_i^c = logAUM_i - \overline{logAUM}.$$

The moderation specification is:

$$Y_{it}^{(k)} = \beta\, ESGUI_{it}^{(12)} + \theta \left(ESGUI_{it}^{(12)} \times logAUM_i^c \right) + \delta\, FR_{it} + \mu_i + \gamma_t + \varepsilon_{it}. \tag{5.2}$$

The interaction coefficient θ shows whether the effect of ESGUI depends on size. If $\theta > 0$, the ESGUI slope is less negative (or more positive) for larger funds, consistent with larger funds having more capacity to maintain disclosure under uncertainty.

Table 5.3 reports the results for the continuous AUM moderation specification, and Figure 5.1 visualises the implied marginal effects evaluated at the 10th, 50th, and 90th percentiles of AUM. Because the interaction is most easily interpreted as a difference in slopes, the discussion focuses on how the estimated ESGUI effect varies with fund size.

The results provide partial support for H2. For *Intensity* (column 1), the interaction coefficient is positive and statistically significant. This indicates that the negative association between ESGUI exposure and the salience of SI disclosure becomes slightly weaker for larger funds. However, the estimated coefficient is extremely small ($\theta \approx 0.000$), implying that the moderation effect is economically negligible. Consistent with this, Figure 5.1 shows only minimal differences in the ESGUI slope across the AUM distribution.

Table 5.3: Moderating Effect of Fund Size (AUM) on ESGUI and SI Measures

	Salience and Breadth				Concreteness / Verifiability
Dep. var.	Intensity	Spectrum	Variety	Scope	Specificity
ESGUI (12m)	-0.001	-0.142*	-0.062*	-0.119*	-0.002
	(0.001)	(0.071)	(0.035)	(0.060)	(0.002)
ESGUI $\times$ log(AUM)	0.000***	0.002	0.005	0.014***	0.001***
	(0.000)	(0.004)	(0.005)	(0.004)	(0.000)
Funding Ratio	-0.005*	-0.087	0.087	-0.149	-0.006
	(0.003)	(0.162)	(0.122)	(0.162)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
Within R^2	0.053	0.009	0.011	0.027	0.027

Notes: Coefficients from OLS regressions with fund and year fixed effects. Standard errors are two-way clustered by fund and publication year-month and shown in parentheses. Within R^2 is the two-way fixed-effects within R^2 (fund and year). *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

For the breadth measures, the evidence is mixed. For *Scope* (column 4), the interaction is positive and statistically significant ($\theta = 0.014$), indicating that larger funds are less likely to reduce the portfolio-wide coverage of SI across asset classes when ESGUI exposure increases. In other words, the contraction in SI breadth documented in Section 5.1 is primarily driven by smaller funds, while larger funds maintain a somewhat broader asset-class coverage under higher ESGUI exposure. By contrast, the interaction terms for *Spectrum* and *Variety* are small and statistically insignificant, suggesting that fund size does not systematically moderate how ESGUI exposure relates to topic breadth or strategy breadth in these specifications.

For *Specificity* (column 5), the interaction coefficient is positive and statistically significant ($\theta = 0.001$). This implies that larger funds are slightly less likely to reduce the concreteness of their SI language when ESGUI exposure increases. However, the magnitude of this moderation effect is very small in economic terms. Overall, the moderation results reinforce the earlier conclusion that the primary adjustment occurs in the salience and breadth of SI disclosure rather than in the concreteness of the language used.

5.3. Moderation by Carbon Exposure (H3)

Hypothesis 3 tests whether the relationship between ESG-related contestation/uncertainty and SI disclosure depends on how carbon-exposed a fund is. I proxy carbon exposure with the weighted average carbon intensity (WACI, Scope 1 + 2). A notable limitation is that WACI is only observed once. I therefore treat WACI as a time-invariant fund characteristic. Because WACI is observed once (2024Q3) and treated as time-invariant, the models in this section do not identify how a fund changes disclosure as it decarbonises. Instead, they test whether more versus less carbon-exposed funds have different disclosure *slopes* with respect to $ESGUI_{it}^{(12)}$.

Let $WACI_i$ denote fund i's carbon exposure and define $logWACI_i = \log(WACI_i)$. For interpretability, I mean-center $logWACI_i$:

$$logWACI_i^c = logWACI_i - \overline{logWACI}.$$

The moderation model extends the baseline specification with an interaction term:

$$Y_{it}^{(k)} = \beta\, ESGUI_{it}^{(12)} + \theta \left(ESGUI_{it}^{(12)} \times logWACI_i^c \right) + \delta\, FR_{it} + \mu_i + \gamma_t + \varepsilon_{it}. \tag{5.3}$$

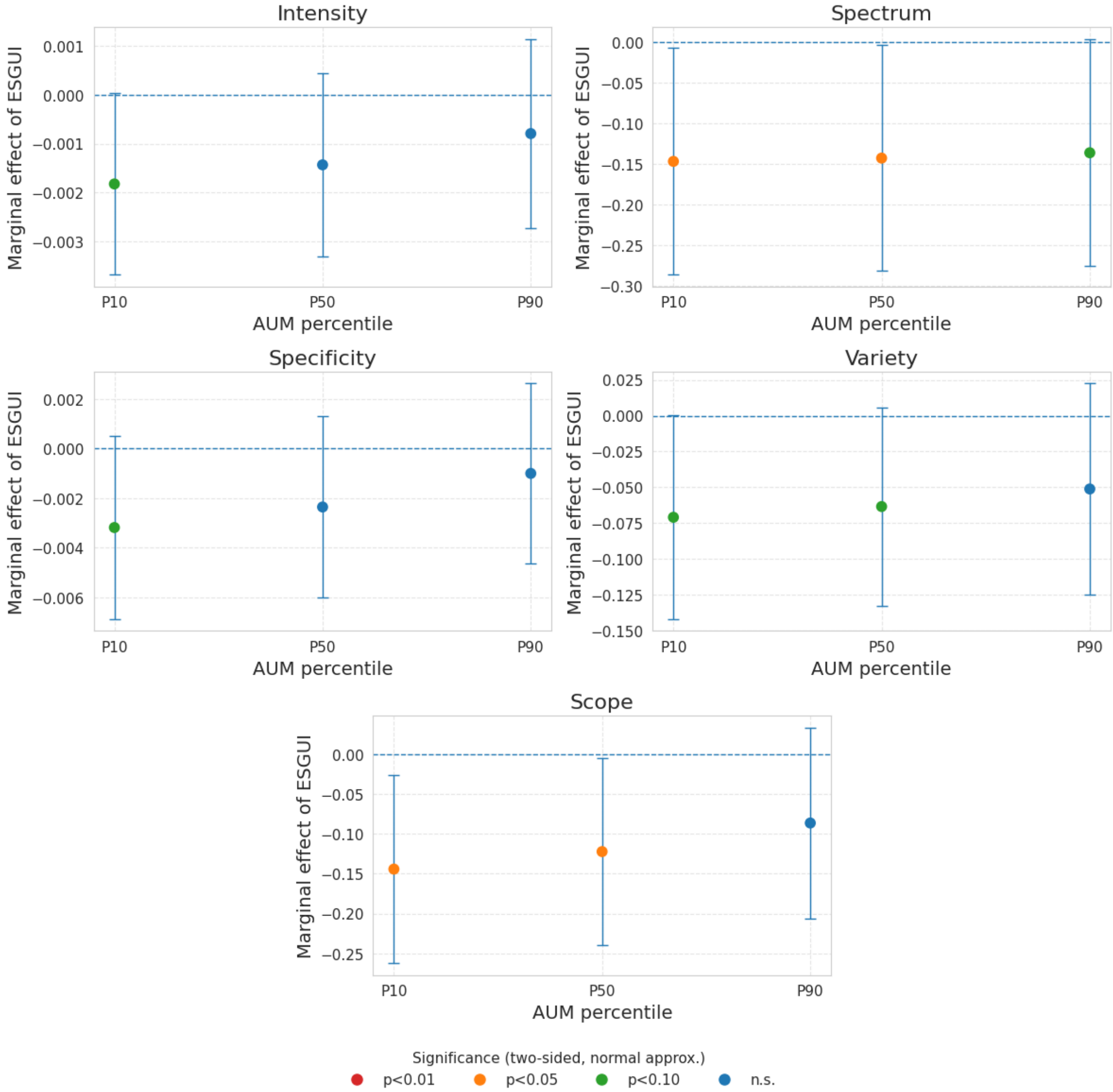

Figure 5.1: Marginal effects of ESGUI on SI measures across fund size. The figure reports estimated marginal effects of ESGUI (12-month rolling mean) evaluated at the 10th, 50th, and 90th percentiles of fund size (AUM). Points indicate marginal effects and whiskers denote 95% confidence intervals. The horizontal dashed line marks zero.

Table 5.4 reports the interaction results across the five SI measures, and Figure 5.2 visualises the implied marginal effects evaluated at the 10th, 50th, and 90th percentiles of WACI.

Table 5.4: Moderating Effect of Carbon Exposure (WACI) on ESGUI and SI Measures

	Salience and Breadth				Concreteness / Verifiability
Dep. var.	Intensity	Spectrum	Variety	Scope	Specificity
ESGUI (12m)	-0.002*	-0.130*	-0.058	-0.152**	-0.003
	(0.001)	(0.075)	(0.037)	(0.069)	(0.002)
ESGUI $\times$ log(WACI)	0.000	-0.021	-0.027***	-0.015	0.000
	(0.000)	(0.013)	(0.007)	(0.012)	(0.000)
Funding Ratio	-0.004	-0.114	0.062	-0.076	-0.005
	(0.002)	(0.170)	(0.118)	(0.164)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	727	727	727	727	727
Within R^2	0.022	0.011	0.020	0.014	0.007

Notes: Coefficients from OLS regressions with fund and year fixed effects. Standard errors are two-way clustered by fund and publication year-month and shown in parentheses. Within R^2 is the two-way fixed-effects within R^2 (fund and year). *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Overall, there is limited evidence that the exposure-disclosure relationship differs strongly by carbon exposure for most outcomes. For *Intensity* (column 1), the interaction coefficient is essentially zero and statistically insignificant, indicating that the negative association between ESGUI exposure and the salience of SI disclosure does not meaningfully differ between more and less carbon-exposed funds.

For *Spectrum* and *Scope* (columns 2 and 4), the interaction terms are also statistically insignificant. The estimated coefficients are negative, which would imply somewhat stronger contractions in topic breadth and asset-class coverage for more carbon-exposed funds, but the estimates are imprecise and do not provide reliable evidence of systematic moderation.

The clearest moderation result appears for *Variety* (column 3). The interaction term $ESGUI \times \log(WACI)$ is negative and statistically significant ($\theta = -0.027$), indicating that the contraction in the number of SI strategies discussed under higher ESGUI exposure is stronger for funds with higher carbon exposure. In other words, while all funds tend to narrow their SI strategy disclosure somewhat when ESG-related uncertainty rises, this narrowing is slightly more pronounced among funds with more carbon-intensive portfolios.

In magnitude terms, moving from the 10th to the 90th percentile of WACI makes the ESGUI slope approximately 0.013 strategy categories more negative. For a one-standard-deviation increase in ESGUI exposure, this corresponds to roughly 0.01 additional SI strategy categories being omitted by more carbon-exposed funds relative to less carbon-exposed funds. Although statistically significant, this moderation effect is economically small and should therefore be interpreted as a subtle difference in disclosure behaviour rather than a substantively large adjustment in reporting practices. Figure 5.2 illustrates this pattern, with the marginal ESGUI effect on *Variety* becoming slightly more negative at higher levels of WACI.

Finally, for *Specificity* (column 5), the interaction coefficient is effectively zero and statistically insignificant, indicating no evidence that the relationship between ESGUI exposure and the concreteness of SI language differs systematically across funds with different carbon exposure.

Taken together, the results provide only limited support for H3. Carbon exposure appears to moderate the ESGUI–disclosure relationship primarily for the strategy-breadth margin (*Variety*), where more carbon-exposed funds reduce the range of SI strategies discussed more strongly under higher ESG-related uncertainty. For the other disclosure dimensions, the estimated moderation effects are small and statistically indistinguishable from zero.

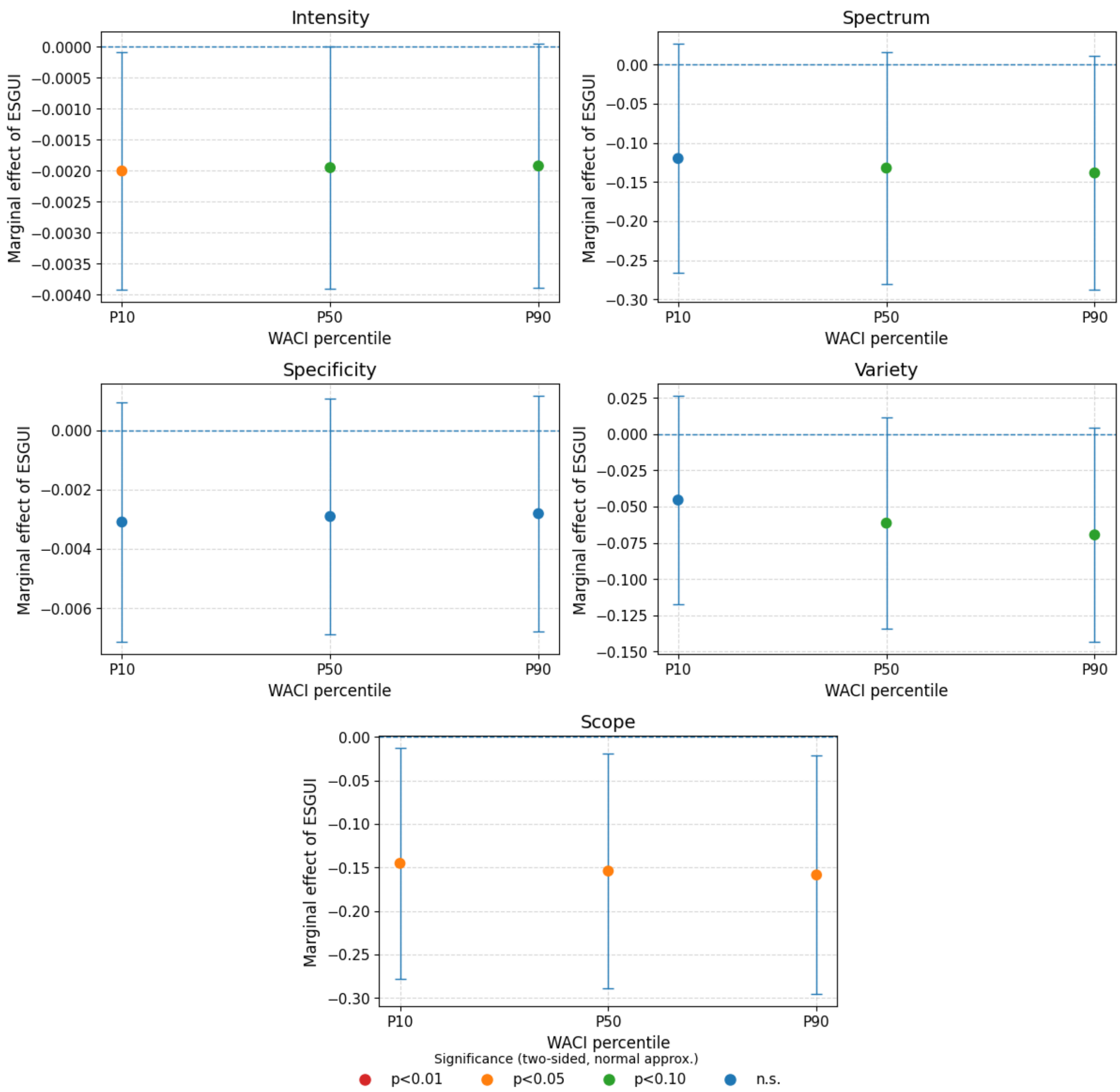

Figure 5.2: Marginal effects of ESGUI on SI measures across carbon exposure (WACI). The figure reports estimated marginal effects of ESGUI (12-month rolling mean) evaluated at the 10th, 50th, and 90th percentiles of funds' carbon exposure (WACI). Points indicate marginal effects and whiskers denote 95% confidence intervals. The horizontal dashed line marks zero.

Joint moderation with fund size

As shown in Figure 4.7, WACI is correlated with fund size. Because fund size is related to reporting capacity and potentially to publication timing, a potential concern is that apparent WACI moderation partly reflects AUM-based differences documented in Section 5.2. To address this possibility, I estimate a joint moderation model that includes interactions of $ESGUI_{it}^{(12)}$ with both $\log(\textit{AUM})$ and $\log(\textit{WACI})$. This specification allows the WACI interaction to be interpreted conditional on fund size and therefore helps assess whether the carbon-exposure moderation reflects carbon intensity in its own right rather than proxying for size-related disclosure capacity. Results from the joint model are reported in Appendix Table A.18.

The main conclusion for carbon exposure remains intact. The negative $ESGUI \times \log(WACI)$ interaction for *Variety* persists with nearly identical magnitude and strong statistical significance ($\theta = -0.025$), indicating that the carbon-exposure moderation for strategic breadth is not driven by fund size. In other words, the tendency for more carbon-exposed funds to narrow the range of SI strategies discussed under higher ESG-related uncertainty remains visible even after accounting for differences in reporting capacity associated with fund size.

Beyond *Variety*, the joint model provides at most weak evidence of carbon-related moderation in *Specificity*. The $ESGUI \times \log(WACI)$ interaction becomes positive and marginally significant, while the baseline ESGUI coefficient remains close to zero. Substantively, this suggests that among more carbon-exposed funds, higher ESG-related uncertainty may be associated with slightly more concrete SI language. However, the magnitude of this effect is small and should be interpreted cautiously.

Figure A.2 plots the implied ESGUI slopes across the WACI distribution while controlling for AUM. The visual pattern closely mirrors the regression results: the negative uncertainty effect on *Variety* becomes somewhat stronger at higher levels of carbon exposure, while the slopes for *Intensity*, *Spectrum*, and *Scope* remain broadly similar across the WACI distribution.

Taken together, the evidence suggests that carbon exposure primarily moderates how uncertainty affects the *diversity of SI strategies* discussed, rather than the overall prominence or portfolio-wide scope of SI disclosure.

5.4. Robustness and Sensitivity Analyses

This section presents several robustness and specification checks to see if the main results depend on seasonal reporting patterns, report length, how the ESGUI exposure window is defined, or certain modeling choices.

One concern is that disclosure practices might change throughout the year. For example, reports published in March could differ from those in October for reasons not related to ESG uncertainty. To address this, I added publication month-of-year fixed effects to the main model. The estimated coefficients stayed similar in sign and, for the main breadth outcomes, in size. The negative links for Variety and Scope remain statistically significant. However, the results for Intensity and Specificity are more sensitive, as their statistical significance changes when month-of-year effects are included (Appendix Table A.12).

Another concern is that the breadth measures might just reflect differences in how long the reports are. Appendix Table A.13 shows that $ESGUI^{(12)}$ does not have a meaningful effect on overall report length. Also, when I control directly for $\log(\textit{total\_sentences})$, the estimated ESGUI effects on Spectrum, Variety, and Scope stay about the same (Table A.14). This supports the idea that the narrowing is selective, not just due to report length.

Next, Appendix Table A.15 checks how results change with different exposure-window definitions. The coefficients are mostly similar in sign, but they are estimated less precisely for shorter windows. This suggests that annual-report disclosure responds more to ongoing uncertainty than to short-term changes.

Finally, Appendix Table A.16 shows functional-form checks using fractional response models for bounded outcomes, and Appendix Table A.17 presents a quadratic specification as a general check for nonlinearity. Overall, these robustness checks support the main finding that higher ESG-related uncertainty is mainly linked to modest contractions in SI disclosure salience and breadth. The evidence for contraction is strongest for *Variety* and *Scope*, generally holds for *Intensity*, and should be interpreted more cautiously for *Spectrum* because of the placebo results.

6

Discussion and Conclusion

Pension funds play an important role in the climate transition because they own significant parts of the economy and invest for the long term. Many have adopted sustainable investing (SI) policies and shared them publicly. However, ESG has become more politically debated and its future is less certain. Expectations can shift quickly, and the reputational and legal risks of making strong ESG statements have grown, even if formal rules stay the same. Since most SI decisions are not visible to outsiders, annual reports offer an important look at how funds approach SI. Still, disclosure is not a neutral act; it is shaped by various pressures. This thesis explores how increases in ESG-related debate and uncertainty in the public discourse affect the way Dutch pension funds report on sustainable investment, and what this means for theory, stakeholders, and how disclosures are designed.

The results suggest a general pattern: when ESG uncertainty is higher in the months preceding publication, SI disclosure tends to become somewhat less extensive and covers fewer areas. Higher exposure to uncertainty is associated with a lower proportion of SI sentences in reports (Intensity) and less coverage across different measures, although the strength of these relationships varies across outcomes. Funds mention fewer types of SI strategies (Variety) and report SI in fewer asset classes (Scope), which are also the dimensions where the results appear most consistent across specifications. There is also some indication that funds talk about fewer SI topics (Spectrum), although this relationship is less robust and should be interpreted with more caution. However, the amount of specific SI statements (Specificity) does not consistently decrease across specifications. Overall, the results suggest that disclosure becomes “less and narrower” rather than simply more vague.

This pattern is consistent with the two response types discussed earlier. In a contraction response, when ESG expectations are uncertain or debated, funds say less about SI and cover fewer aspects, which reduces the risk of criticism and avoids making the report seem like a firm commitment. In a hedging response, funds continue to discuss SI but use less specific language to stay flexible. Overall, the evidence is more consistent with a contraction response than with hedging. The largest and most consistent changes appear in how much and how widely SI is discussed, rather than in making the language more vague. At the same time, the estimated effects are economically modest, indicating gradual adjustments in disclosure focus rather than large shifts in reporting behaviour.

Contraction also has a practical meaning. Lower Intensity means SI receives somewhat less attention compared to other parts of the report. Lower breadth means SI is covered in fewer topics, strategy types, and asset classes. The analysis in this thesis is based on when reports are published: after accounting for fund and year differences, the main variation comes from funds publishing in different months, which links them to different levels of national ESG uncertainty. This provides a realistic but limited source of variation. The results should therefore be interpreted as steady changes in focus and coverage rather than sudden shifts in behaviour that occur when uncertainty increases.

The lack of a consistent change in Specificity further clarifies what contraction means in practice. When funds continue to discuss SI, they typically do not remove concrete details such as targets, metrics, or deadlines. This makes the hedging explanation less convincing. One reason may be that parts of annual reports follow templates or established disclosure standards that require a minimum level of detail. Another reason is that reducing the amount and range of SI content may be easier than rewriting existing statements to be less specific, especially if reporting teams want to avoid inconsistencies across documents. Finally, disclosure is not the same as action: less extensive SI reporting can mean that funds continue their practices but choose not to highlight them publicly when ESG debates are heated.

These results can also be understood by looking at previous research on sustainability disclosure and how organisations respond to outside pressure. Environmental accounting studies have long shown that sustainability reporting is influenced by concerns about legitimacy. For instance, Patten (2002) explains that organisations change their environmental disclosures when they face more scrutiny or shifting public expectations, using reporting to manage their reputation. Reviews of ESG disclosure research find similar patterns: sustainability reporting often involves strategic communication choices instead of just neutral descriptions of what organisations do (Velte, 2023). The pattern of contraction found in this thesis fits with this view. When debates about ESG become more uncertain or controversial, pension funds may narrow their sustainability claims to avoid criticism or being seen as inconsistent. However, the relatively small size of the effects found here suggests that these changes are gradual adjustments in emphasis, not major shifts in reporting.

The findings also align with research on how organisations react to uncertainty. Studies on economic policy uncertainty show that companies act more cautiously when future policies are harder to predict (Baker et al., 2016). While most of that research looks at investment and finance decisions, the same idea can apply to public communication. When it is unclear what to expect from regulations or politics, organisations may avoid making strong promises or broad public statements. In sustainability reporting, this caution may be even stronger because of increasing legal and reputational risks around ESG statements. As Osovsky (2015) notes, organisations are facing more questions about whether their sustainability claims can be backed up. If the meaning and measurement of ESG concepts are also debated (Edmans, 2023), limiting the range of public claims may be a way to protect themselves. The results in this thesis suggest that pension funds show this kind of behaviour in their annual reports, though the strength of the effect differs depending on the type of disclosure.

Differences between funds help explain who adjusts disclosure the most. Smaller funds show some indication of a somewhat larger drop, particularly in how prominently they highlight SI and how broadly they present it across their portfolios. This is consistent with the idea that smaller organisations have fewer resources to monitor policy debates, manage communication strategies, and maintain extensive reporting processes. Larger funds, by contrast, may have greater internal capacity and may face stronger expectations from supervisors, media, and beneficiaries to maintain transparent reporting. Nevertheless, the estimated differences in responsiveness between funds remain modest and should be interpreted cautiously.

Carbon exposure offers another perspective on reputational risk. Using weighted-average carbon intensity (WACI) as a measure of carbon exposure, the most notable difference appears in Variety: funds with higher carbon exposure reduce the range of SI strategies they mention more strongly when ESG uncertainty is high. One possible interpretation is that these funds face greater scrutiny and therefore adopt a more cautious communication strategy. Listing many sustainability strategies in a politically contested environment can raise expectations and attract criticism from different groups. Reducing the number of strategies highlighted may therefore be a way to manage reputational exposure. For most other measures, however, the effect of carbon exposure is smaller and less clear, suggesting that it mainly influences which strategies are emphasised rather than driving a broad reduction in disclosure.

The robustness tests generally support the contraction interpretation. Alternative model specifications, including approaches that account for the bounded nature of some outcomes, produce results that are similar to the baseline estimates. Timing tests also show that the relationship between uncertainty and disclosure is strongest when uncertainty is measured over a period that aligns with the annual reporting cycle, particularly twelve months. This pattern supports the idea that annual reports reflect longer-term reporting environments rather than short-term fluctuations. Controlling for report length does not eliminate the negative relationships in breadth measures, suggesting that the results are not simply driven by shorter reports.

At the same time, the identification strategy has clear limitations. Because ESG uncertainty is measured at the national level, the analysis relies on differences in report publication timing across funds within the same year. While this provides useful variation, it also means that the results should be interpreted as suggestive associations rather than precise causal estimates. The falsification test further highlights this point: the relationship between ESG uncertainty and Spectrum remains significant even when uncertainty is measured after publication, which weakens the timing logic for that measure. As a result, the strongest evidence for contraction comes from Variety and Scope, with supportive but less robust evidence for Intensity.

These findings contribute to the literature by linking ESG debate and uncertainty to how pension funds communicate sustainable investing, rather than focusing on asset prices or company investment decisions. The thesis also introduces a "two-margin" view of disclosure behaviour: organisations can either reduce the amount and scope of sustainability communication (contraction) or maintain disclosure but reduce specificity (hedging). In this setting, contraction appears to be the more common response. Methodologically, the thesis also demonstrates how text-based indicators can capture multiple dimensions of sustainability disclosure and how modern language models can help construct such indicators efficiently.

Beyond academic contributions, the results have practical relevance for financial supervision. Pension funds in the Netherlands operate under increasing expectations to report on sustainability risks and responsible investment practices, and a large part of this reporting occurs in narrative form in annual reports. The findings suggest that the breadth and emphasis of such disclosures may change when ESG debates intensify, even if underlying investment practices remain stable. For supervisors such as De Nederlandsche Bank (DNB), this implies that changes in narrative reporting should be interpreted carefully during periods of heightened public debate.

At the same time, the methods used in this thesis illustrate how text-based analysis could support supervisory monitoring. By tracking changes in disclosure focus, scope, and specificity across funds and over time, supervisors could identify shifts in reporting behaviour that may signal changing perceptions of regulatory or reputational risk. Such indicators could complement existing supervisory reviews of pension fund reporting. A set of practical recommendations on how supervisors such as DNB could use similar tools to monitor sustainability disclosure and reduce reporting uncertainty is presented in Appendix A.4.

More broadly, the findings illustrate a classic challenge for supervision. When expectations are unclear and reputational risks increase, organisations may become more cautious in public communication even if underlying practices remain unchanged. In this sense, narrative disclosure functions as a kind of public good: it improves transparency and comparability but may decline precisely when uncertainty increases. Supervisory communication, guidance, and consistent reporting expectations can therefore play an important role in maintaining the usefulness of sustainability disclosures during periods of intense debate.

Several limitations should be acknowledged. Ideally, research on sustainable investing would observe both detailed portfolio decisions and the internal processes through which pension funds develop and communicate their SI strategies. With such data, it would be possible to test whether changes in disclosure correspond to changes in investment behaviour, engagement activities, or voting practices. In practice, however, most of these decisions are not publicly observable. As a result, this thesis relies on annual report disclosures as a visible proxy for how funds communicate their sustainability approach.

The measure of ESG uncertainty used in the analysis also has limitations. ESGUI combines ESG-related attention and uncertainty language in news coverage and therefore does not capture pure policy uncertainty. The analysis cannot fully distinguish whether the observed effects are driven by increased debate about ESG more broadly or by uncertainty about regulatory developments. In addition, the carbon exposure measure is time-invariant in the available data, which makes it possible to compare funds with different carbon profiles but does not allow changes in decarbonisation strategies to be tracked over time.

These limitations point to several promising directions for future research. First, improved data linking narrative disclosures with portfolio holdings, engagement records, or voting behaviour could help determine whether contraction in reporting reflects "quiet ESG" communication or genuine changes in investment strategy. Second, future research could explore how reporting teams interpret ESG debates and translate them into communication decisions, for example through interviews or qualitative case studies. Third, testing similar hypotheses in other institutional contexts would help assess the generality of the findings. Asset managers, insurance companies, and listed firms operate under different reporting norms and regulatory pressures, which may shape how sustainability disclosure responds to uncertainty. Cross-country comparisons could further reveal how national regulatory environments influence these dynamics.

To conclude, when Dutch pension funds face more debate and uncertainty about ESG, they tend to disclose somewhat less and cover fewer sustainable investment topics, although these changes are generally modest in size. The clearest evidence appears in reductions in the breadth of strategies and asset classes discussed, with weaker or less consistent effects for other dimensions. These findings suggest that narrative sustainability disclosure is not only a tool for transparency but can also function as a strategic communication instrument that contracts when reputational or political risks increase. For both beneficiaries and supervisors, this highlights the importance of maintaining comparable and informative sustainability reporting, particularly during periods when the public debate around ESG becomes more intense. *When the debate gets louder and funds grow quieter, that is exactly when supervisors should listen more closely.*

References

Aerts, Walter and Denis Cormier (2009). “Media legitimacy and corporate environmental communication”. In: *Accounting, Organizations and Society* 34.1, pp. 1–27. URL: https://EconPapers.repec.org/RePEc:eee:aosoci:v:34:y:2009:i:1:p:1-27.

Agnese, Paolo, Fabio De Masi, Paolo Porretta, and Federico Santoboni (2024). “Sustainable or not sustainable pension fund: This is the question. The case of environmental social governance policies in the Italian pension system”. In: *Socio-Economic Planning Sciences* 94, p. 101954. DOI: 10.1016/j.seps.2024.101954. URL: https://doi.org/10.1016/j.seps.2024.101954.

Arayakarnkul, P., P. Chatjuthamard, and S. Treepongkarun (2022). “Board gender diversity, corporate social commitment and sustainability”. In: *Corporate Social Responsibility and Environmental Management* 29.5, pp. 1706–1721. DOI: 10.1002/csr.2320. URL: https://doi.org/10.1002/csr.2320.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis (2016). “Measuring Economic Policy Uncertainty”. In: *The Quarterly Journal of Economics* 131.4, pp. 1593–1636. DOI: 10.1093/qje/qjw024.

Bams, Dennis, Peter C. Schotman, and Mukul Tyagi (Mar. 2016). *Asset Allocation Dynamics of Pension Funds*. Netspar Discussion Paper 03/2016-016. Available at SSRN. Netspar, p. 35. DOI: 10.2139/ssrn.2766491. URL: https://ssrn.com/abstract=2766491.

Bauer, Rob, Rien Bogman, Matteo Bonetti, and Dirk Broeders (2020). *The Impact of Trustees’ Age and Representation on Strategic Asset Allocations*. Tech. rep. Working Paper No. 698. De Nederlandsche Bank. URL: https://www.dnb.nl/media/mqgkvsjb/working_paper_698.pdf.

Bauer, Rob, Dirk Broeders, and Annick van Ool (2023). “Walk the green talk? A textual analysis of pension funds’ disclosures of sustainable investing”. In: *Journal of Pension Economics & Finance* 24.2, pp. 297–325. DOI: 10.1017/S147474722400009X. URL: https://doi.org/10.1017/S147474722400009X.

Bauer, Rob, Tobias Ruof, and Paul Smeets (2021). “Get Real! Individuals Prefer More Sustainable Investments”. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3287430. URL: https://papers-ssrn-com.tudelft.idm.oclc.org/sol3/papers.cfm?abstract_id=3287430.

Bingler, Julia Anna, Mathias Kraus, Markus Leippold, and Nicolas Webersinke (2022). “Cheap talk and cherry-picking: What ClimateBERT has to say on corporate climate risk disclosures”. In: *Finance Research Letters* 47, p. 102776. DOI: 10.1016/j.frl.2022.102776. URL: https://doi.org/10.1016/j.frl.2022.102776.

BNP Paribas (Feb. 2024). *Political resistance to ESG concerns Dutch institutional investors*. Insights — Sustainability & CSR. BNP Paribas Netherlands. URL: https://www.bnpparibas.nl/en/political-resistance-to-esg-concerns-dutch-institutional-investors/ (visited on 11/03/2025).

Brammer, S. and S. Pavelin (2006). “Voluntary environmental disclosures by large UK companies”. In: *Journal of International Financial Management & Accounting* 17.4, pp. 467–495. DOI: 10.1111/j.1468-5957.2006.00598.x.

Buda, Mateusz, Atsuto Maki, and Maciej A. Mazurowski (2018). “A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks”. In: *Neural Networks* 106, pp. 249–259. DOI: 10.1016/j.neunet.2018.07.011. URL: https://doi.org/10.1016/j.neunet.2018.07.011.

Cho, Charles H., Matias Laine, Robin W. Roberts, and Michelle Rodrigue (2015). “Organized hypocrisy, organizational façades, and sustainability reporting”. In: *Accounting, Organizations and Society* 40.1, pp. 78–94. DOI: 10.1016/j.aos.2014.12.003. URL: https://doi.org/10.1016/j.aos.2014.12.003.

Cho, Charles H. and Dennis M. Patten (2007). “The role of environmental disclosures as tools of legitimacy: A research note”. In: *Accounting, Organizations and Society* 32.7-8, pp. 639–647. DOI: 10.1016/j.aos.2006.09.009.

Christensen, Hans B., Luzi Hail, and Christian Leuz (2021). “Mandatory CSR and sustainability reporting: Economic analysis and literature review”. In: *Review of Accounting Studies* 26.3, pp. 1176–1248. DOI: 10.1007/s11142-021-09609-5. URL: https://doi.org/10.1007/s11142-021-09609-5.

Cifrino, David (2023). *The Politicization of ESG Investing*. URL: https://www.sir.advancedleadership.harvard.edu/articles/politicization-of-esg-investing.

De Nederlandsche Bank (Apr. 2018). *Proportional and Effective Supervision*. Tech. rep. De Nederlandsche Bank. URL: https://www.dnb.nl/media/vo2nldcq/proportional-and-effective-supervision.pdf.

— (2025). *Pensioenfondsen (macro-economisch) Dashboard*. https://www.dnb.nl/statistieken/dashboards/pensioenfondsen/pensioenfondsen-macro-economisch/. Accessed: Tuesday 10th March, 2026.

Delmas, Magali A. and Vanessa C. Burbano (2011). "The Drivers of Greenwashing". In: *California Management Review* 54.1, pp. 64–87. DOI: 10.1525/cmr.2011.54.1.64.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv* abs/1810.04805. DOI: 10.48550/arXiv.1810.04805. URL: https://arxiv.org/abs/1810.04805.

Dohle, Mona (2024). *Dutch pension funds in the firing line over ESG performance*. URL: https://netzeroinvestor.net/news-and-views/dutch-pension-funds-in-the-firing-line-over-esg-performance-link.

Du, Wei, Laksh Advani, Yashmeet Gambhir, Daniel Perry, Prashant Shiralkar, Zhengzheng Xing, and Aaron Colak (Dec. 2023). "Effective Proxy for Human Labeling: Ensemble Disagreement Scores in Large Language Models for Industrial NLP". In: *Proceedings of the Third Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*. Singapore: Association for Computational Linguistics, pp. 53–61. URL: https://aclanthology.org/2023.gem-1.5/.

Edmans, Alex (2023). "The End of ESG". In: *Financial Management* 52.1, pp. 3–17. DOI: 10.1111/fima.12413. URL: https://doi.org/10.1111/fima.12413.

European Commission (2024). *Sustainability-related disclosure in the financial services sector*. Overview page on the Sustainable Finance Disclosure Regulation (SFDR) and related acts. URL: https://finance.ec.europa.eu/sustainable-finance/disclosures/sustainability-related-disclosure-financial-services-sector_en (visited on 10/16/2025).

European Securities and Markets Authority (June 2024). *Final Report on Greenwashing: Response to the European Commission's request for input on "greenwashing risks and the supervision of sustainable finance policies"*. Tech. rep. ESMA36-287652198-2699. European Securities and Markets Authority. URL: https://www.esma.europa.eu/sites/default/files/2024-06/ESMA36-287652198-2699_Final_Report_on_Greenwashing.pdf.

Fabian, Frederick, Claris Parenti, Maddy Taylor, and Valerio Micale (Dec. 2025). *State of OECD Pension Funds' Climate Transition: Insights and Recommendations from the Net Zero Finance Tracker*. Climate Policy Initiative. URL: https://www.climatepolicyinitiative.org/publication/state-of-oecd-pension-funds-climate-transition-insights-and-recommendations-from-the-net-zero-finance-tracker/.

Fouche, Gwladys (Nov. 4, 2025). "Norway poised to pause wealth fund's ethical divestments". In: *Reuters*. URL: https://www.reuters.com/business/finance/norway-poised-pause-wealth-funds-ethical-divestments-2025-11-04/.

Frénay, Benoît and Michel Verleysen (2014). "Classification in the Presence of Label Noise: A Survey". In: *IEEE Transactions on Neural Networks and Learning Systems* 25.5, pp. 845–869. DOI: 10.1109/TNNLS.2013.2292894. URL: https://doi.org/10.1109/TNNLS.2013.2292894.

Friede, Gunnar, Timo Busch, and Alexander Bassen (2015). "ESG and financial performance: aggregated evidence from more than 2000 empirical studies". In: *Journal of Sustainable Finance & Investment* 5.4, pp. 210–233. DOI: 10.1080/20430795.2015.1118917. URL: https://doi.org/10.1080/20430795.2015.1118917.

Gavriilidis, Konstantinos (2021). "Measuring Climate Policy Uncertainty". In: DOI: 10.2139/ssrn.3847388. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3847388.

Golka, Philipp and Natascha van der Zwan (2025). "Asymmetric autonomy: pension fund investing between members and markets". In: *Socio-Economic Review*, mwaf042. DOI: 10.1093/ser/mwaf042. URL: https://doi.org/10.1093/ser/mwaf042.

Gulen, Huseyin and Mihai Ion (2016). "Policy Uncertainty and Corporate Investment". In: *The Review of Financial Studies* 29.3, pp. 523–564. DOI: 10.1093/rfs/hhv050. URL: https://doi.org/10.1093/rfs/hhv050.

Gull, Ammar Ali, Nazim Hussain, Sana□Akbar Khan, Zaheer Khan, and Asif Saeed (2023). "Governing Corporate Social Responsibility Decoupling: The Effect of the Governance Committee on Corporate Social Responsibility Decoupling". In: *Journal of Business Ethics* 185.2, pp. 349–374. DOI: 10.1007/s10551-022-05181-3. URL: https://doi.org/10.1007/s10551-022-05181-3.

Hahn, Rüdiger and Michael Kühnen (2013). "Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research". In: *Journal of Cleaner Production* 59, pp. 5–21. DOI: 10.1016/j.jclepro.2013.07.005.

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (2019). "Firm-level political risk: Measurement and effects". In: *The Quarterly Journal of Economics* 134.4, pp. 2135–2202. DOI: 10.1093/qje/qjz021.

Healy, Paul M. and Krishna G. Palepu (2001). "Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature". In: *Journal of Accounting & Economics* 31.1–3, pp. 405–440. DOI: 10.1016/S0165-4101(01)00018-0.

Heseltine, Michael and Bernhard Clemm von Hohenberg (2024). "Large language models as a substitute for human experts in annotating political text". In: *Research & Politics* 11.1, p. 20531680241236239. DOI: 10.1177/20531680241236239. URL: https://journals.sagepub.com/doi/10.1177/20531680241236239.

Jens, Candace E. (2017). "Political Uncertainty and Investment: Causal Evidence from U.S. Gubernatorial Elections". In: *Journal of Financial Economics* 124.3, pp. 563–579. DOI: 10.1016/j.jfineco.2016.01.034. URL: https://doi.org/10.1016/j.jfineco.2016.01.034.

Johnson, Justin M. and Taghi M. Khoshgoftaar (2019). "Survey on Deep Learning with Class Imbalance". In: *Journal of Big Data* 6.1, pp. 1–54. DOI: 10.1186/s40537-019-0192-5. URL: https://doi.org/10.1186/s40537-019-0192-5.

Krueger, Philipp, Zacharias Sautner, Dragon Yongjun Tang, and Rui Zhong (2023). *The Effects of Mandatory ESG Disclosure around the World*. Tech. rep. European Corporate Governance Institute (ECGI). URL: https://www.ecgi.global/sites/default/files/working_papers/documents/kreugersautnertangzhongfinalfeb_0.pdf.

Leuz, Christian and Peter D. Wysocki (2016). *The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research*. Tech. rep. European Corporate Governance Institute (ECGI) – Law Working Paper No. 306/2016. DOI: 10.2139/ssrn.2733831. URL: https://ssrn.com/abstract=2733831.

Li, Victor and Ani Nenkova (2015). "Fast and Accurate Prediction of Sentence Specificity". In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI Press, pp. 665–671. DOI: 10.1609/aaai.v29i1.9517. URL: https://doi.org/10.1609/aaai.v29i1.9517.

Lu, Yao, Max Bartolo, Andrew Moore, Sebastian Riedel, and Pontus Stenetorp (2022). "Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity". In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 8086–8098. URL: https://aclanthology.org/2022.acl-long.558.

Luan, Tian (2024). "A Review of Corporate Social Responsibility Decoupling and Its Impact: Evidence from China". In: *Sustainability* 16.10, p. 4047. DOI: 10.3390/su16104047. URL: https://doi.org/10.3390/su16104047.

Lyon, Thomas P. and John W. Maxwell (Mar. 2006). *Greenwash: Corporate Environmental Disclosure Under Threat of Audit*. Tech. rep. Working Paper No. 1055. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=938988. Ross School of Business, University of Michigan / Kelley School of Business, Indiana University.

Marotta, Fulvia, Maria Sole Pagliari, and Jasper de Winter (2025). *Uncertainty and Climate Policy in the Netherlands: Measure and Economic Effects*. DNB Analysis. Views are those of the authors and do not necessarily reflect DNB positions. Amsterdam: De Nederlandsche Bank.

Marquis, Christopher, Michael W. Toffel, and Yanhua Zhou (2016). "Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing". In: *Organization Science* 27.2, pp. 483–504. DOI: 10.1287/orsc.2015.1039. URL: https://doi.org/10.1287/orsc.2015.1039.

McDonnell, C. (2024). "Pension funds and fossil fuel phase-out: historical developments and limitations of pension climate strategies". In: *International Environmental Agreements: Politics, Law and Economics* 24, pp. 169–191. DOI: 10.1007/s10784-024-09626-0. URL: https://doi.org/10.1007/s10784-024-09626-0.

McHugh, Mary L. (2012). “Interrater reliability: the kappa statistic”. In: *Biochemia Medica* 22.3, pp. 276–282. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/.

Muñoz, David Ramos, Marco Lamandini, and Michele Siri (2024). *The implementation of the Sustainability-related Financial Disclosures Regulation (SFDR)*. Tech. rep. PE 754.212. European Parliament. URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2024/754212/IPOL_STU(2024)754212_EN.pdf.

NOS (Nov. 4, 2025). “Noors oliefonds zet ethisch investeringsbeleid in de ijskast”. nl. In: *NOS Nieuws*. URL: https://nos.nl/artikel/2589353-noors-oliefonds-zet-ethisch-investeringsbeleid-in-de-ijskast.

Olasehinde-Williams, Godwin, Oktay Özkan, and Seyi Saint Akadiri (2023). “Effects of climate policy uncertainty on sustainable investment: A dynamic analysis for the U.S.” In: *Environmental Science and Pollution Research* 30.19, pp. 55326–55339. DOI: 10.1007/s11356-023-26257-1. URL: https://doi.org/10.1007/s11356-023-26257-1.

Ongan, Serdar, Ismet Gocer, and Cem Işık (2025). “Introducing the New ESG-Based Sustainability Uncertainty Index (ESGUI)”. In: *Sustainable Development* 33.3. Accepted Author Manuscript available via Strathprints, pp. 4457–4467. DOI: 10.1002/sd.3351. URL: https://onlinelibrary.wiley.com/doi/10.1002/sd.3351.

Osovsky, Adi (2015). *Puffery on the Market: A Behavioral Economic Analysis of the Puffery Defense in the Securities Arena*. Tech. rep. SSRN. DOI: 10.2139/ssrn.2603051. URL: https://ssrn.com/abstract=2603051.

Papke, Leslie E. and Jeffrey M. Wooldridge (1996). “Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates”. In: *Journal of Applied Econometrics* 11.6, pp. 619–632. DOI: 10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1.

Pastor, Luboš and Pietro Veronesi (2011). *Political Uncertainty and Risk Premia*. Tech. rep. Working Paper No. 17464. Cambridge, MA: National Bureau of Economic Research. DOI: 10.3386/w17464. URL: https://www.nber.org/papers/w17464.

Patten, Dennis M. (2002). “The relation between environmental performance and environmental disclosure: A research note”. In: *Accounting, Organizations and Society* 27.8, pp. 763–773. DOI: 10.1016/S0361-3682(02)00028-4. URL: https://doi.org/10.1016/S0361-3682(02)00028-4.

Pensions UK (2024). *Majority of pension funds have net zero commitment but challenges persist in meeting climate and nature targets*. URL: https://www.pensionsuk.org.uk/News/Article/Majority-of-pension-funds-have-net-zero-commitment-but-challenges-persist-in-meeting-climate-and-nature-targets-.

PensionsEurope (2023). *IORP II Directive*. Accessed: 2025-09-30. URL: https://pensionseurope.eu/policy-priorities/iorp-ii-directive/.

Ravina, Enrichetta and Nicola Persico (2025). “ESG Choice with Polarized Investors”. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.5086547. URL: https://papers-ssrn-com.tudelft.idm.oclc.org/sol3/papers.cfm?abstract_id=5086547.

Al-Shaer, Habiba and Mahbub Zaman (2016). “Board gender diversity and sustainability reporting quality”. In: *Journal of Contemporary Accounting & Economics* 12.3, pp. 210–222. DOI: 10.1016/j.jcae.2016.09.001. URL: https://doi.org/10.1016/j.jcae.2016.09.001.

Shaikh, Imlak (2022). “On the relationship between policy uncertainty and sustainable investing”. In: *Journal of Modelling in Management* 17.4, pp. 1504–1523. DOI: 10.1108/JM2-12-2020-0320.

TCFD (2023). *Task Force on Climate-related Financial Disclosures*. Financial Stability Board. URL: https://www.fsb-tcfd.org/ (visited on 10/16/2025).

UNFCCC (2015). *The Paris Agreement*. https://unfccc.int/process-and-meetings/the-paris-agreement. Accessed: 2025-10-27.

Vasquez, Omar (2023). “Unbundling ESG for Fiduciary Integrity: An Empirical Analysis on the Representativeness of ESG Pension Investment Legislation”. In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4776762. URL: https://papers-ssrn-com.tudelft.idm.oclc.org/sol3/papers.cfm?abstract_id=4776762.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017). “Attention is All You Need”. In: URL: https://arxiv.org/pdf/1706.03762.pdf.

Velte, Peter (2023). “Determinants and Financial Consequences of Environmental Performance and Reporting: A Literature Review of European Archival Research”. In: *Journal of Environmental Management* 340, p. 117916. DOI: 10.1016/j.jenvman.2023.117916.

Verrecchia, Robert E. (2001). “Essays on Disclosure”. In: *Journal of Accounting and Economics* 32.1–3, pp. 97–180. DOI: 10.1016/S0165-4101(01)00025-8. URL: https://doi.org/10.1016/S0165-4101(01)00025-8.

Werken aan ons Pensioen (n.d.). *Overzicht pensioenfondsen, pensioenverzekeraars en premiepensioeninstellingen*. https://www.werkenaanonspensioen.nl/belangrijke-begrippen/overzicht-pensioenfondsen-en-pensioenverzekeraars. Accessed: Tuesday 10th March, 2026. URL: https://www.werkenaanonspensioen.nl/belangrijke-begrippen/overzicht-pensioenfondsen-en-pensioenverzekeraars.

Zhang, Chiyuan, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals (2017). “Understanding Deep Learning Requires Rethinking Generalization”. In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=Sy8gdB9xx.

A

Appendix

A.1. SI Classification: Construction and Validation

To compute SI disclosure measures from the reports, SI-related sentences are first isolated. This is done with a two-stage pipeline: (i) a transparent, high-recall rule-based filter for candidate flagging and traceability, and (ii) a fine-tuned transformer classifier that produces the final SI labels. Restricting analysis to the SI subset enables direct computation of SI intensity (the share of SI-related content) and supports additional SI-specific quality indicators on the relevant sentence set.

Each sentence is first flagged as potentially SI-related using broad English and Dutch keyword patterns and heuristics. The rules search for signals such as ESG and sustainability terminology, thematic issues (e.g. climate change, human rights), investment context, and references to regulatory or reporting frameworks. The predicted rule label and the specific triggering rule are stored alongside each sentence to ensure traceability. This stage is used as a diagnostic layer and to maximise recall.

Final SI classification is performed with RobBERT, a Dutch RoBERTa-based transformer language model. Transformer models represent text by learning relationships between all words in a sentence simultaneously, rather than sequentially[7]. RobBERT belongs to the BERT family[8]. BERT-style models are pre-trained on large unlabelled corpora and then fine-tuned on labelled examples for task-specific classification. This makes them suitable for scalable sentence-level topic classification. The choice follows Bauer, Broeders, et al. (2023), who also fine-tune RobBERT for SI sentence classification.

Whereas Bauer, Broeders, et al. (2023) hand-label 2,000 sentences, this thesis evaluates whether LLM-assisted annotation can reduce human labelling effort. Two approaches are compared: (i) prompt-based Agentic labelling and (ii) prompt-based API labelling. The resulting curated dataset is then used to fine-tune RobBERT.

Agentic labelling. In the Agentic setup, ChatGPT 5.1 Agent mode labels a batch of sentences supplied as a single CSV, using a fixed definition and decision rules (Appendix A.1). Agent mode can maintain context across the batch, perform internal consistency checks, and occasionally revise earlier labels after seeing later cases. This may increase within-batch coherence but can introduce dependence across sentences and potential order effects: labels may vary with input ordering and the agent's self-review trajectory in a given run, consistent with evidence that LLM outputs can be sensitive to ordering (Lu et al., 2022). To assess accuracy, I hand-labelled > 500 sentences and targeted a 2,000-sentence training set.

Run-to-run variability and aggregation. Applying the Agentic procedure three times to the same 535

[7] A transformer model is a type of neural language model that represents text by learning relationships between all words in a sentence simultaneously, rather than processing them sequentially (Vaswani et al., 2017).

[8] BERT (Bidirectional Encoder Representations from Transformers) is a neural language model that learns contextual word representations by processing text bidirectionally, meaning it considers both left and right context simultaneously. It is pre-trained on large unlabeled corpora using tasks such as masked language modelling and next sentence prediction, and can then be fine-tuned for downstream classification tasks with relatively limited labelled data (Devlin et al., 2019).

hand-labelled sentences produces substantial variation across runs (Table A.1). A simple majority vote yields high overall performance (97.0% accuracy; F1 = 0.936), but heterogeneous run quality implies that weak runs can dominate the aggregated decision on borderline cases. For training data, this is problematic because discriminative models can be sensitive to systematic label noise.

Table A.1: Performance metrics per agent en majority vote (535 cases)

Metric	Agent 1	Agent 2	Agent 3	Majority vote
Total cases	535	535	535	535
Correctly classified	92.3%	97.4%	83.7%	97.0%
TP / FP / TN / FN	99 / 12 / 395 / 29	121 / 7 / 400 / 7	107 / 66 / 341 / 21	117 / 5 / 402 / 11
Precision	0.892	0.945	0.618	0.959
Recall (sensitivity)	0.773	0.945	0.836	0.914
Specificity	0.971	0.983	0.838	0.988
Balanced accuracy	0.872	0.964	0.837	0.951
F1 score	0.828	0.945	0.711	0.936
Cohen's κ (vs. user)	0.779	0.928	0.601	0.917

Disagreement-based review and curated labels. Following Heseltine et al. (2024) and Du et al. (2023), I therefore use two independent Agentic passes with the same prompt and manually adjudicate only the disagreement cases. Evaluation then focuses on residual errors where both runs agree but are wrong. This yields near-best accuracy (balanced accuracy 0.972) and strong precision/specificity, with high inter-rater reliability (Table A.2; $\kappa = 0.950$)[9].

Table A.2: Performance Comparison: Agent 1, Agent 2, and the Consensus cases

Metric	Agent 1	Agent 2	Consensus
Total cases	535	535	497
Correctly classified	92.3%	97.4%	96.4%
TP / FP / TN / FN	99 / 12 / 395 / 29	121 / 7 / 400 / 7	97 / 3 / 392 / 5
Precision	0.892	0.945	0.970
Recall (sensitivity)	0.773	0.945	0.951
Specificity	0.971	0.983	0.992
Balanced accuracy	0.872	0.964	0.972
F1 score	0.828	0.945	0.960
Cohen's κ (vs. user)	0.779	0.928	0.950

API labelling benchmark. As a benchmark, I also label sentences via API using ChatGPT 4o-mini[10]. Each sentence is classified independently, which improves reproducibility. Three API runs on the same 535 sentences show high stability and similar performance (Table A.3). All runs use temperature $=$ 0 to minimise sampling-related variability. However, disagreement-based review provides little gain because the runs share highly consistent errors, producing very few disagreements to correct (Table A.4).

The final training set is built using two independent Agentic annotation runs combined with targeted human adjudication. Agent mode achieves high performance but exhibits non-trivial run-to-run variation and potential order effects, making a single pass less reliable. The API approach is more reproducible but shows largely systematic, shared errors and therefore offers limited scope for disagreement-based

[9] Cohen's kappa is a statistical measure of agreement between two annotators that accounts for agreement occurring by chance. Unlike simple percentage agreement, it adjusts for the fact that annotators may sometimes assign the same label randomly. Values range from −1 to 1, where higher values indicate stronger agreement, with values above zero reflecting agreement beyond chance (McHugh, 2012).

[10] An application programming interface (API) is a standardised way for software to communicate with another service. In this thesis, "using an LLM through an API" means sending text inputs (here: individual sentences and a fixed prompt) to a model hosted by a provider and receiving the model's output (here: a binary label) programmatically, enabling automated and reproducible batch processing.

Table A.3: SI classification performance metrics per API run and majority vote

Metric	API run 1	API run 2	API run 3	Majority vote
Total cases	535	535	535	535
Correctly classified	95.9%	95.7%	95.5%	95.9%
TP / FP / TN / FN	115 / 9 / 398 / 13	114 / 9 / 398 / 14	114 / 10 / 397 / 14	115 / 9 / 398 / 13
Precision	0.927	0.927	0.919	0.927
Recall (sensitivity)	0.898	0.891	0.891	0.898
Specificity	0.978	0.978	0.975	0.978
Balanced accuracy	0.938	0.934	0.933	0.938
F1 score	0.913	0.908	0.905	0.913
Cohen's κ (vs. user)	0.886	0.880	0.875	0.886

Table A.4: Performance Comparison: API run 1, API run 2, and the Consensus cases.

Metric	API run 1	API run 2	Consensus
Total cases	535	535	532
Correctly classified	95.9%	95.7%	96.1%
TP / FP / TN / FN	115 / 9 / 398 / 13	114 / 9 / 398 / 14	114 / 8 / 397 / 13
Precision	0.927	0.927	0.934
Recall (sensitivity)	0.898	0.891	0.898
Specificity	0.978	0.978	0.980
Balanced accuracy	0.938	0.934	0.939
F1 score	0.913	0.908	0.916
Cohen's κ (vs. user)	0.886	0.880	0.890

correction. The two-run Agentic workflow balances scalability with label quality: labels are accepted when both runs agree, and only disagreements are reviewed manually.

RobBERT is fine-tuned to classify sentences as SI-related vs. not SI-related using the curated Agentic labels. Because the data are imbalanced (fewer SI sentences), training uses a class-weighted loss to penalise minority-class errors more heavily. Sentences are tokenised with a maximum sequence length of 64 tokens[11]. Evaluation uses a stratified train–test split to preserve the SI/non-SI class proportions in both sets. Optimisation targets the F1-score rather than accuracy, since accuracy can be misleading in imbalanced settings.

Table A.5: Classification Report: SI-Intensity Classifier (RoBERTa), using curated Agentic labelled data

Class	Precision	Recall	F1-score	Support
Not SI-related	0.977	0.993	0.985	301
SI-related	0.979	0.929	0.953	99
Accuracy			0.978	400
Macro avg	0.978	0.961	0.969	400
Weighted avg	0.978	0.978	0.977	400

On the held-out stratified test set ($n = 400$), the model achieves 0.978 accuracy and strong class-specific performance (Table A.5). For the SI class ($n = 99$), precision is 0.979 and recall is 0.929 (F1 = 0.953), indicating few false positives and high retrieval of SI-related sentences. Because the test labels follow the curated annotation scheme (two Agentic runs with adjudicated disagreements), the reported metrics primarily reflect the model's ability to reproduce that scheme; external validity may depend on

[11]Tokenisation is the process by which a language model converts text into smaller units, called tokens, that can be processed numerically. These tokens may correspond to whole words, parts of words, or punctuation, depending on the model's vocabulary. Tokenisation allows the model to handle text of varying length and structure in a consistent way. In practice, one token corresponds on average to approximately four characters of common English text, which is roughly three quarters of a word.

alignment with independent human judgements in new domains or styles.

SI Classification Prompt

Shown below is the exact prompt used to classify sentences as SI-related (1) or not SI-related (0). The same prompt was used (i) in ChatGPT 5.1 Agent mode for batch annotation and (ii) in API calls for per-sentence labelling.

```
Instructie voor LLM-classificatie van zinnen over Sustainable Investing (SI)

Taak
Je ontvangt individuele zinnen afkomstig uit documenten van Nederlandse pensioenfondsen.
Taak: Label elke zin als:
   1 -> SI-gerelateerd
   0 -> Niet SI-gerelateerd

Een zin is SI-gerelateerd wanneer deze functioneel gaat over duurzaam beleggen,
maatschappelijk verantwoord beleggen, ESG-integratie of stewardship.

Je output dus per zin slechts één getal: 1 of 0.

Definitie van Sustainable Investing (SI)
SI omvat o.a.:
   - ESG-integratie (environmental, social, governance)
   - Maatschappelijk verantwoord beleggen (MVB)
   - Engagement/stewardship gericht op duurzaamheid
   - Uitsluitingscriteria (tabak, wapens, fossiel e.d.)
   - Duurzame transitie (klimaat, CO2-reductie, net-zero)
   - Themabeleggingen (klimaat, biodiversiteit, sociale standaarden)
   - Impactbeleggen (mits gekoppeld aan beleggingen)
   - Regelgeving en transparantie (SFDR, EU-taxonomie, CSRD)

Geef label 1 wanneer:
De zin functioneel betrekking heeft op:
   - Beleggingsbeleid rondom duurzaamheid
   - "duurzaam beleggen", "maatschappelijk verantwoord beleggen"
   - ESG-integratie in besluitvorming
   - MVB-beleid, implementatie, evaluatie
   - SI-activiteiten
   - engagement met bedrijven over ESG/doelen
   - stewardship
   - uitsluitingen
   - impactfondsen (mits belegging-gerelateerd)
   - Regulering en rapportage
   - SFDR / CSRD / EU-taxonomie
   - Artikel 8/9 mandaat
   - metingen, rapportage, monitoring van ESG
   - Onderwerpen gekoppeld aan beleggingen:
   - klimaatrisico's
   - CO2-footprint
   - energietransitie
   - mensenrechtenpraktijken waar het fonds invloed op heeft

Voorbeelden van label 1:
   - "Het fonds integreert ESG-factoren in het beleggingsproces."
   - "Wij sluiten wapens uit volgens het MVB-beleid."
   - "Onze portefeuille is afgestemd op de EU-taxonomie."
   - "Engagement richt zich op CO2-reductie en mensenrechten."

Geef label 0 wanneer:
De zin niet over beleggingen EN duurzaamheid gaat.
Dat omvat zinnen over:
```

```
   - financieel rendement zonder duurzaamheid
   - intern bestuur (vergaderingen, commissies) zonder ESG-koppeling
   - actuariële, juridische of operationele processen
   - communicatie, rapportage, compliance zonder duurzaamheidscontext
   - personeelszaken, HR, IT, organisatiezaken

Voorbeelden van label 0:
   - "Het rendement bedroeg 7%."
   - "De bestuursvergadering vond plaats op 4 maart."
   - "Het fonds voldoet aan EMIR-richtlijnen."
   - "De portefeuille werd geherbalanceerd vanwege marktrisico."

Twijfelgevallen -- Beslissingsregels
1) "Maatschappelijk" of "verantwoord" -> Alleen 1 als beleggingsgerelateerd
   - 1: "Maatschappelijk verantwoord beleggen is beleid."
   - 0: "Maatschappelijk verantwoord beloningsbeleid."
2) Governance
   - 1 wanneer governance betrekking heeft op duurzaam beleggen
   - 0 wanneer governance algemeen/administratief is
3) Sociale of ecologische thema's
   - 1 wanneer gekoppeld aan beleggingen/engagement
   - 0 wanneer maatschappelijk probleem zonder beleggingscontext
4) "Impact"
   - 1: impactbeleggingen / engagement-impact
   - 0: impact in algemene zin of auditingimpact

Woorden die zelf geen SI impliceren:
Deze woorden triggeren niet automatisch SI, tenzij gekoppeld aan beleggen/beleid:
   - impact
   - maatschappelijk
   - verantwoord
   - ESG (zonder context)
   - communicatie
   - governance
   - rapportage
   - monitoring
   - stemrecht

Beslisstrategie (in volgorde):
Is de zin functioneel over duurzaam beleggen, beleid of engagement?
   -> Label 1
Gaat de zin alleen over organisatie/financien/governance zonder ESG-koppeling?
   -> Label 0
Twijfel: bevat duurzaamheidstermen maar zonder SI-context?
   -> Label 0
Twijfel: bevat beleggen maar geen duurzaamheid?
   -> Label 0

Consistentie
Pas dezelfde beslisregels toe op elke zin.
Structurele fout is erger dan fout op één geval.

Output per zin
Alleen:
   0 of 1

Mini-checklist voordat je labelt
Beantwoord voor jezelf:
Heeft dit invloed op duurzame beleggingskeuzes?
   Ja -> 1
   Nee -> 0
```

A.2. Specificity Classification: Construction and Validation

The second task concerns the *specificity* of SI-related sentences. SI detection identifies *what* topics are discussed, but not *how informative* the disclosure is. In sustainability reporting, funds may discuss SI in broad, aspirational terms without disclosing concrete actions, targets, timelines, or measurable outcomes. A specificity classifier therefore distinguishes substantive, decision-relevant disclosure from boilerplate language and enables a quality measure beyond volume or intensity.

This dimension is also used by Bauer, Broeders, et al. (2023), who evaluate the informativeness of sustainable investing statements by separating general rhetoric from concrete, verifiable disclosure. Unlike Bauer, Broeders, et al. (2023), this thesis classifies *sentences* rather than paragraphs and constructs a *specificity ratio* (the fraction of SI sentences that are specific), rather than a count of specific paragraphs.

Sentence-level classification offers a more reliable segmentation for annual reports: paragraph boundaries in PDF extractions are often inconsistent due to layout, hyphenation, and parsing artefacts, while sentences provide stable units that match transformer models' strengths. The ratio-based measure is also better aligned with the econometric hypotheses (H1–H3), which predict proportional shifts in disclosure precision under ESG-policy uncertainty. Counts are mechanically affected by report length and formatting; ratios capture disclosure composition. Because ratios can be unstable when SI content is sparse, I impose a minimum-SI-sentence threshold and, in robustness checks, downweight fund–years with very low SI sentence counts.

Specificity classification is substantially harder than SI detection. SI detection largely relies on topical cues (e.g. ESG, climate, emissions, social issues), which are often explicit and lexical. Specificity instead concerns *concreteness and verifiability*: whether a sentence contains a measurable, time-bound, or otherwise externally checkable commitment, action, or performance attributable to the fund. The linguistic literature similarly emphasises that specificity depends on semantic and contextual properties rather than surface form alone (Li et al., 2015). Examples of specific SI sentences and the underlying mechanisms are shown in Table A.6.

Table A.6: Examples of specific SI-related disclosure statements (translated from Dutch to English from annual reports)

Specificity mechanism	Example sentence
Measurable target (A)	"It is our ambition to reduce the CO_2 emissions of all investments by 50% by 2030, measured relative to 2019."
Time-bound quantitative goal (A+B)	"The primary objective formulated is a reduction of CO_2 emissions from the investment portfolio, with the goal of having reduced total portfolio emissions by 25% in 2022 compared to the first measurement in 2018."
Concrete exclusion action (C)	"In 2018, ["The Fund"] decided to exclude companies involved in the production of thermal coal, such as Adani Enterprises, Arch Coal, and Coal India."
Concrete engagement action with scope (C)	"In 2023, we initiated the next phase of the energy transition engagement programme, shifting the focus to 50 major consumers of fossil fuels and raw materials within our portfolio."
Reporting / procedural obligation (D)	"The PAI report for 2024 must be published by 30 June 2025 and included in a section entitled 'Sustainability Information' on the pension fund's website."
Quantified performance outcome (A)	"In 2024, the carbon footprint (Scopes 1 and 2) of our investment portfolio was 49% lower than in the base year."
Explicit measurement methodology (D)	"["The Fund"] monitors ESG risks using, among other tools, the ESG meter, and since 2021 has measured both the carbon footprint and carbon intensity."
Portfolio-level allocation constraint (A+C)	"For mortgage funds, a requirement was introduced in 2023 that at least 80% of the assets invested in mortgages must be allocated to funds classified as Article 8 under the SFDR framework."

While LLMs can make these distinctions, converting them into exhaustive and consistently applied rules is difficult: borderline cases are common when sentences are evaluated in isolation. In practice, both humans and models may rely on correlates of specificity (numbers, time markers, named actions, explicit commitments) even when instructed to apply a more semantic definition. This creates systematic risks: aspirational statements with a date or number can be over-labelled as specific, while concrete actions expressed without explicit metrics can be missed. Fully automated labelling also struggles to incorporate attribution and discourse context (e.g. fund action versus background, regulation, or third-party activity). These issues motivate careful prompt design, targeted human review, and explicit documentation of residual error patterns (Li et al., 2015).

Agentic batch classification is unstable for specificity: repeating the same task in Agent mode produces substantial run-to-run variation, consistent with shifting interpretations of borderline cases during within-batch self-review. Because reproducibility is essential for constructing reference labels, and because performance fluctuated too strongly to justify relying on a single Agentic run, I conclude that Agent mode is not a reliable labelling strategy for SI-specificity. Instead, I use independent, per-sentence API classification as the primary labelling approach.

Specificity is rare: within the SI corpus, less than 10% of sentences appear specific. A purely random gold sample would therefore contain too few positives to estimate minority-class performance reliably. I therefore construct a $n = 500$ gold set that combines random sampling with targeted oversampling of likely positives.

To create the targeted pool transparently and with high recall, I first flag “likely specific” candidates using regular expressions. This step is not intended to classify, only to increase the yield of potential positives for manual labelling. I then apply minor refinements to remove clear non-cases (e.g. purely definitional or contextual sentences without fund-attributable action). Applied to a subset of 3,000 SI sentences, this produces a candidate pool of 757 sentences.

The final gold set draws 350/500 sentences randomly from the full SI pool (to preserve naturalistic distribution) and 150/500 from the candidate pool (to increase positive support). All 500 sentences are then hand-labelled using the same definition and decision rules as in the main protocol.[12] Because the gold set over-represents positives (about 20%), I report class-conditional metrics (precision, recall, F1, balanced accuracy) and interpret prevalence-dependent metrics with reference to the randomly sampled portion.

Table A.7 reports performance for three repeated API runs and their majority vote against the gold labels. Performance is highly consistent across runs: overall accuracy is about 91% (90.8–91.2%), and Cohen’s κ (0.717–0.729) indicates substantial agreement but also confirms that specificity is notably harder than SI detection.

Table A.7: Performance metrics per API run and majority vote against gold labels ($n = 500$)

Metric	API run 1	API run 2	API run 3	Majority vote
Total cases	500	500	500	500
Correctly classified	90.8%	90.8%	91.2%	91.0%
TP / FP / TN / FN	79 / 19 / 375 / 27	80 / 20 / 374 / 26	80 / 18 / 376 / 26	80 / 19 / 375 / 26
Precision	0.806	0.800	0.816	0.808
Recall (sensitivity)	0.745	0.755	0.755	0.755
Specificity	0.952	0.949	0.954	0.952
Balanced accuracy	0.849	0.852	0.855	0.853
F1 score	0.775	0.777	0.784	0.780
Cohen’s κ (vs. gold)	0.717	0.719	0.729	0.724

Minority-class performance is moderate: precision is around 0.80–0.82 (18–20 false positives), while recall is lower at about 0.75 (26–27 false negatives). The main limitation is therefore sensitivity to specific statements rather than filtering non-specific ones; negative-class specificity is high (0.949–0.954). Majority voting does not improve results, implying that errors are strongly correlated across runs under a fixed prompt and deterministic decoding. Likewise, a disagreement-based “consensus” approach yields no meaningful gains because the best runs disagree on very few cases (Table A.8).

The remaining errors are systematic. Manual inspection shows false positives often arise when the

[12] The prompt and decision rules used for specificity labelling are reported in the corresponding prompt appendix.

model treats formal cues (a year, a number, policy terminology) as sufficient for specificity, even when the sentence is aspirational or contextual (visions, planned reviews, framework descriptions, generic stewardship references). False negatives occur when concrete, verifiable actions are expressed without the cues the model appears to prioritise: implementation changes, engagement actions, and methodological disclosure about monitoring and measuring indicators (including scopes). Overall, the model follows a stable but simplified decision rule that overweights easily detectable surface features and underweights agency and verifiability.

I experimented with multiple refinements of the specificity prompt in an attempt to improve performance. Although these adjustments changed the model's decision boundary, they did not yield a consistent improvement in the evaluation metrics and did not recover the stronger results reported in Table A.8. At the same time, the API outputs remained highly consistent across repeated runs, indicating that stochasticity is not the main driver of error. Instead, the remaining disagreements are largely systematic, which is consistent with the inherent ambiguity of specificity judgements and with the practical limits of perfectly consistent hand coding by a single annotator. Performance should therefore be interpreted relative to human reliability: the objective is to obtain a robust and replicable operational measure of SI-specificity, rather than perfect agreement with one individual coder on borderline cases. For this reason, and given the reproducibility of the API-based approach, I use API classification to construct the specificity-labelled training set for the subsequent modelling stages.

Table A.8: Performance Comparison: API run 1, API run 2, and the Consensus cases.

Metric	**API run 1**	**API run 2**	**Consensus**
Total cases	500	500	498
Correctly classified	90.8%	91.2%	91.2%
TP / FP / TN / FN	79 / 19 / 375 / 27	80 / 18 / 376 / 26	79 / 18 / 376 / 26
Precision	0.806	0.816	0.814
Recall (sensitivity)	0.745	0.755	0.752
Specificity	0.952	0.954	0.954
Balanced accuracy	0.849	0.855	0.853
F1 score	0.775	0.784	0.782
Cohen's κ (vs. gold)	0.717	0.729	0.727

Given the run-to-run stability in Tables A.7–A.8, I treat the API procedure as a reliable operational labeller and apply it to the unlabeled corpus. Because performance may vary across funds and writing styles, I interpret these outputs as "silver" labels with stable measurement error rather than as error-free ground truth.

To build a training set that balances learning capacity and robustness, I use targeted class balancing. Specific sentences are rare, so purely random sampling would yield too few positives. At the same time, indiscriminately adding large volumes of automatically labelled data can amplify label noise, to which high-capacity discriminative models are sensitive (Buda et al., 2018; Johnson et al., 2019; Zhang et al., 2017; Frénay et al., 2014). I therefore (i) apply the API classifier to 4,000 SI sentences and retain all predicted specific sentences (465 positives), and (ii) sample a set of predicted non-specific sentences uniformly at random as negatives to preserve stylistic and topical diversity.

To check generalisation beyond the gold set, I manually reviewed a subsample of 150/465 predicted positives. Among these, 23 are false positives, implying an estimated precision of about 85% in this subsample. This aligns with the gold-set precision (about 0.80–0.82) and suggests that the targeted sampling procedure yields a reasonably accurate, though not noise-free, positive set. The false positives are largely systematic: they often discuss exclusion, engagement, voting, or climate objectives at a general policy level without verifiable detail, and they frequently mention timelines or evaluations without measurable indicators, thresholds, or explicit scope. The resulting labels are therefore best understood as an operational SI-specificity measure with known, stable error, suitable for comparative and econometric use when interpreted accordingly.

Table A.9: Classification report for specificity classifier (test set)

Class	Precision	Recall	F1-score	Support
Not specific	0.907	0.930	0.918	157
Specific	0.876	0.839	0.857	93
Accuracy			0.896	250
Macro average	0.892	0.884	0.888	250
Weighted average	0.896	0.896	0.896	250

The results in Table A.9 indicate strong performance on both classes (accuracy 89.6%). Precision is high and balanced, while recall is somewhat lower for the “specific” class (0.839), indicating that some truly specific statements are missed. The F1-score of 0.857 for the specific class suggests that the model captures the main patterns of specificity, with remaining ambiguity concentrated in borderline cases.

Specificity Classification Prompt

Shown below is the exact prompt used to classify sentences as Specific (1) or not Specific (0). The same prompt was used (i) in ChatGPT 5.1 Agent mode for batch annotation and (ii) in API calls for per-sentence labelling.

```
PROMPT VOOR LLM-AGENT -CLASSIFICATIE VAN SPECIFICITEIT

Je krijgt één zin uit een jaarverslag van een Nederlands pensioenfonds.
Deze zin is al eerder geclassificeerd als gerelateerd aan duurzaam of maatschappelijk
    verantwoord beleggen (SI).

Jouw taak

Bepaal of de zin specifiek of niet-specifiek is volgens de onderstaande formele criteria.
Je geeft uitsluitend één van de twee labels terug:
   - 1 →Specifiek
   - 0 →Niet-specifiek
Niets anders.

1. Definitie: Wat is een specifieke SI-zin?
Een zin is specifiek wanneer zij een verifieerbare, concrete of meetbare toezegging, actie of
    prestatie bevat die door het fonds zelf wordt uitgevoerd of geclaimd.
Een zin is specifiek wanneer zij ten minste één van de volgende elementen bevat:

(A) Meetbare doelstelling of KPI
   - percentages, aantallen, reductiedoelen
   - emissiemetingen (bijv. CO-intensiteit, WACI)
   - beleggingsallocaties in exacte hoeveelheden

(B) Tijdgebonden elementen
   - duidelijke doeljaren "(in "2030, "voor "2025, ""jaarlijks)
   - deadlines of periodes

(C) Concrete, afgebakende handelingen
   - uitsluitingen of desinvesteringen in een specifieke sector of groep
   - engagement-acties met duidelijke doelgroep (bedrijven, sectoren)
   - stembeleid met concrete procedure of doel
   - toepassing van ESG-criteria op een benoemd deel van de portefeuille

(D) Procedurele of rapportage-verplichtingen
   - terugkerende, toetsbare acties "(wij rapporteren …"jaarlijks)
   - expliciete methodologieën "(wij meten emissies volgens …")
```

```
Kernprincipe:
De zin moet op zichzelf een element bevatten dat extern controleerbaar of falsifieerbaar is.

2. Wat is niet-specifiek?
Een zin is niet-specifiek wanneer zij:
    - vage, algemene of aspiratieve taal bevat zonder meetbare component
    - algemene ESG-waarden, intenties of filosofieën beschrijft
    - beleid beschrijft zonder concreet fonds-specifiek element
    - regelgeving bespreekt zonder implementatie door het fonds
    - geen getallen, targets, tijdsbegrenzing of afgebakende acties bevat

Voorbeelden van niet-specifiek:
    - "Wij vinden duurzaamheid belangrijk".
    - "Wij streven naar een betere wereld".
    - "Wij integreren ESG in ons beleggingsproces". (zonder details)
    - "Volgens SFDR moeten …"fondsen (zonder uitleg wat het fonds doet)

3. Edge-cases: beslisregels
De zin verwijst naar een actie, maar zonder details →0
    - "Wij zetten ons actief in voor engagement". →niet-specifiek
    - De zin bevat een getal, maar het is niet SI-gerelateerd →0
    - "Wij beheren €12 miljard aan vermogen". →geen SI-specificiteit
De zin bevat een tijdswoord zonder concrete actie →0
    - "De komende jaren blijven wij verduurzamen". →niet-specifiek
De zin beschrijft een concrete actie maar zonder scope →1 als het extern controleerbaar is
    - "Wij hebben bedrijven aangesproken op hun emissies". →specifiek (actie is controleerbaar)

De zin bevat meerdere acties, maar één is specifiek →1

Alleen één verifieerbare component is voldoende.

Incomplete bijzinnen

De zin telt alleen als specifiek als in diezelfde zin de verifieerbare component aanwezig is.

4. Voorbeelden
Specifiek (1)
    - "Wij reduceren onze  CO-intensiteit met 50% in "2030.
    - "In 2023 voerden we 1620 engagementgesprekken met hoog-emitterende bedrijven".
    - "Wij sluiten kolenmijnbedrijven uit die meer dan 20% van hun omzet uit kolen halen".
    - "Wij rapporteren jaarlijks onze WACI-score volgens PCAF-richtlijnen".

Niet-specifiek (0)
    - "Duurzaamheid is een belangrijk onderdeel van ons beleid".
    - "Wij streven naar een klimaatneutrale toekomst".
    - "Wij nemen ESG-factoren mee in ons beleggingsproces".
    - "Volgens SFDR worden nieuwe regels ingevoerd".

5. Outputformaat
Je antwoord is exact één teken:
    - 1 als de zin specifiek is
    - 0 als de zin niet-specifiek is

Geen uitleg, geen extra tekst, geen JSON, geen toelichting.
PROMPT EINDE
```

A.3. Additional Results and Robustness Tests

This appendix reports additional analyses supporting the H1 results. The material serves two purposes. First, several *diagnostic checks* document the identifying variation and probe assumptions underlying the publication-aligned design, such as whether publication timing responds systematically to ESG-related uncertainty. Second, a set of *robustness and sensitivity checks* evaluates whether the main findings depend on modelling choices, seasonal reporting patterns, or alternative constructions of the ESGUI exposure measure.

Timing endogeneity test

Table A.10 reports two diagnostics on whether publication timing responds systematically to ESG uncertainty. Column (1) uses the deviation of a fund's publication month from its fund-specific median month, and shows no statistically significant association with $ESGUI^{(12)}$. Column (2) uses an indicator for switching publication month relative to the previous year and yields a small, marginally significant positive estimate ($p < 0.10$), suggesting that higher uncertainty may be associated with a slightly higher likelihood of month switching.

Overall, the diagnostics provide limited evidence of timing responses, motivating robustness checks that control for publication-month seasonality and assess whether results are sensitive to excluding observations with unstable timing.

Table A.10: Association Between ESG Uncertainty and Publication Timing

Dep. var:	(1) Dev. from median pub. month	(2) Switch publication month
ESGUI (12m mean)	0.191	0.041*
	(0.125)	(0.023)
Funding Ratio	-0.017	-0.005
	(0.086)	(0.038)
Fund FE	Yes	Yes
Report-year FE	Yes	Yes
Observations	781	781
Within R^2	0.071	0.010

Notes: Column (1) uses as dependent variable the deviation of a fund's publication month from its fund-specific *median* publication month (median is used for robustness to occasional delays and because months are discrete). Column (2) uses an indicator equal to 1 if a fund publishes in a different calendar month than in the previous reporting year, and 0 otherwise (linear probability model). Both specifications include fund and report-year fixed effects. Standard errors are two-way clustered by fund and publication year-month and are shown in parentheses. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-group variation explained by the included regressors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Remaining variation for identification

Table A.11 and Figure A.1 provide descriptive information on the identifying variation used in the main regressions. All variables are residualised with respect to fund fixed effects, report-year fixed effects, and the baseline controls. The residualised ESG uncertainty measure has a standard deviation of 0.85, showing that still some within-year variation in publication-aligned uncertainty remains after removing common report-year effects and time-invariant fund characteristics.

Figure A.1 plots the partial relationships between residualised ESG uncertainty and residualised SI disclosure measures. These plots are descriptive and do not constitute independent tests, but they help visualise the within-fund, within-year associations that underpin the fixed-effects estimates.

Table A.11: Descriptive Statistics for Residualised Variables

	count	mean	std	min	25%	50%	75%	max
Residual ESGUI (12m)	781	0.000	0.851	-4.655	-0.264	-0.050	0.303	10.000
Residual Intensity	781	0.000	0.016	-0.083	-0.009	0.000	0.010	0.128
Residual Spectrum	781	0.000	1.440	-6.067	-0.891	-0.027	0.935	5.391
Residual Variety	781	0.000	0.819	-2.757	-0.461	0.029	0.503	2.686
Residual Scope	781	0.000	1.276	-4.128	-0.778	0.008	0.841	4.221
Residual Specificity	781	0.000	0.043	-0.135	-0.026	-0.002	0.021	0.325

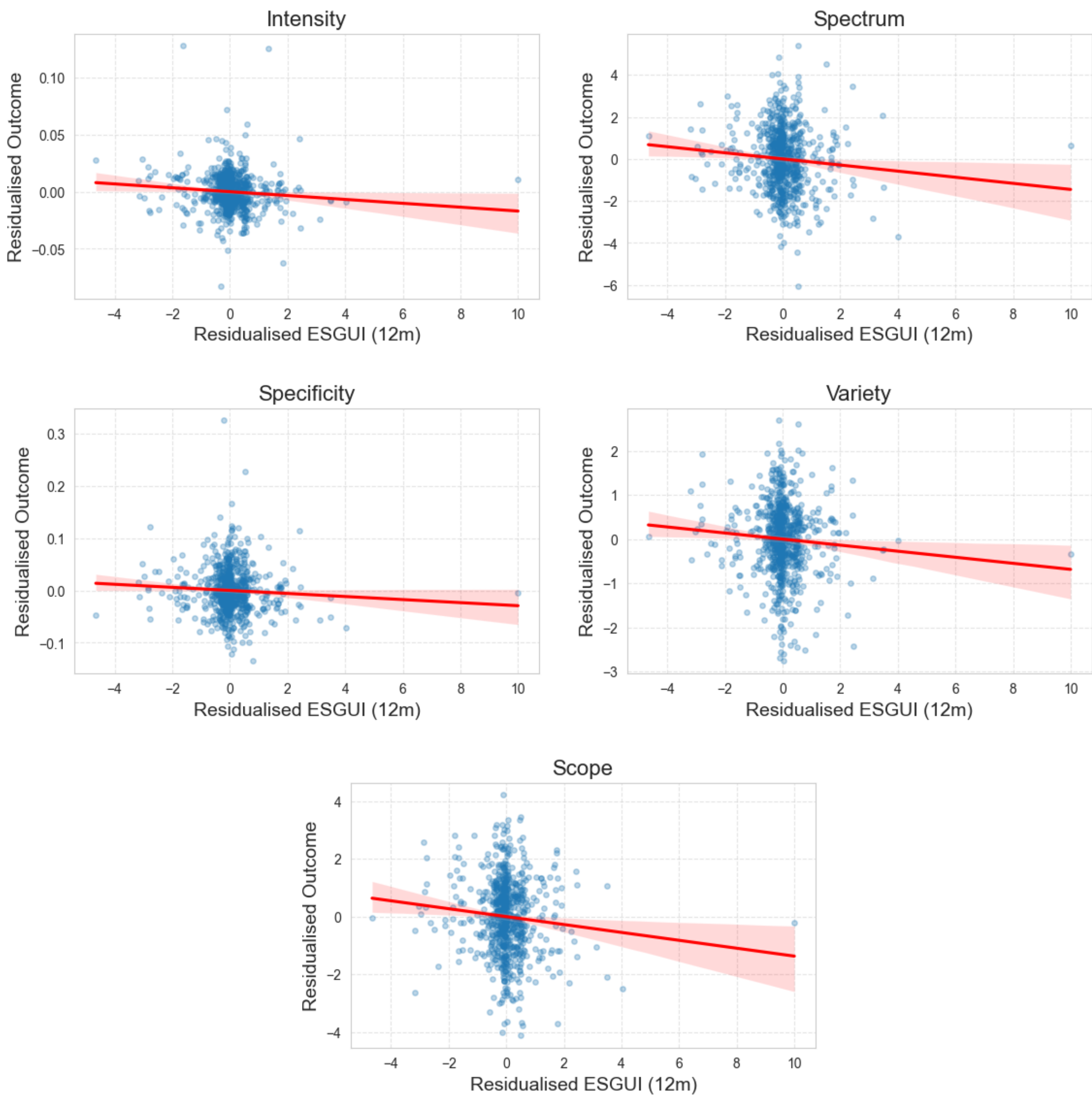

Figure A.1: Residualised ESG policy uncertainty and residualised SI disclosure measures. Each panel plots the relationship between residualised ESG policy uncertainty (12-month rolling window) and residualised SI disclosure outcomes. All variables are residualised with respect to fund fixed effects, report-year fixed effects, and control variables. Points represent fund–year observations; solid lines show linear fitted values with 95% confidence intervals.

Seasonality robustness test

To verify that the baseline relationships are not driven by seasonal reporting-cycle patterns, I augment the baseline specification with publication month-of-year fixed effects. These controls absorb systematic differences in disclosure associated with the timing of report publication within the calendar year.

Table A.12 compares the baseline specification with fund and report-year fixed effects (Panel A) to the augmented model including month-of-year fixed effects (Panel B). The estimated ESGUI coefficients remain broadly similar in sign and magnitude across specifications. In particular, the negative associations for *Variety* and *Scope* remain statistically significant, indicating that the contraction in SI disclosure breadth is not driven by seasonal reporting patterns.

Some sensitivity appears for *Intensity* and *Specificity*, where statistical significance changes across specifications. Overall, however, the results confirm that the main contraction findings are not an artifact of seasonal publication timing.

Table A.12: Baseline Regression Results With and Without Publication Month-of-Year Fixed Effects

Dep. var:	(1) Intensity	(2) Spectrum	(3) Variety	(4) Scope	(5) Specificity
Panel A: Fund and report-year fixed effects					
ESGUI (12m)	-0.002*	-0.145**	-0.069*	-0.137**	-0.003
	(0.001)	(0.071)	(0.036)	(0.062)	(0.002)
Funding Ratio	-0.005*	-0.088	0.085	-0.156	-0.006
	(0.003)	(0.160)	(0.122)	(0.165)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Report-year FE	Yes	Yes	Yes	Yes	Yes
Publication month-of-year FE	No	No	No	No	No
Observations	781	781	781	781	781
Within R^2	0.022	0.008	0.006	0.011	0.007
Panel B: Fund, report-year, and publication month-of-year fixed effects					
ESGUI (12m)	-0.002	-0.097*	-0.067*	-0.145**	-0.004**
	(0.001)	(0.054)	(0.036)	(0.070)	(0.002)
Funding Ratio	-0.005*	-0.103	0.073	-0.164	-0.006
	(0.003)	(0.163)	(0.121)	(0.164)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Report-year FE	Yes	Yes	Yes	Yes	Yes
Publication month-of-year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
Within R^2	0.020	0.004	0.005	0.011	0.009

Notes: Panel A reports coefficients from OLS regressions with fund fixed effects and report-year fixed effects. Panel B adds publication month-of-year fixed effects. Standard errors (in parentheses) are two-way clustered by fund and publication year-month. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-fund variation explained by the included regressors. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Controlling for report length

A potential concern is that the baseline results for the breadth measures reflect mechanical differences in report verbosity rather than substantive changes in disclosure. Longer reports naturally contain more sentences and may therefore mention a larger number of SI topics, strategies, or asset classes. If ESG-related uncertainty were associated with systematically shorter reports, the observed contraction in breadth could arise mechanically.

Table A.13: Association Between ESG Uncertainty and Annual Report Length

Dep. Var:	$\log($*total_sentences*$)$
ESGUI (12m mean)	-0.001
	(0.006)
Funding Ratio	-0.013
	(0.012)
Fund FE	Yes
Report-year FE	Yes
Observations	781
Within R^2	0.004

Notes: The dependent variable is $\log($*total_sentences*$)$. The regression includes fund and report-year fixed effects. Standard errors are two-way clustered by fund and publication year-month and are shown in parentheses. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-group variation explained by the included regressors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To examine this possibility, Table A.13 tests whether ESG policy uncertainty is related to overall report length. The dependent variable is $\log($*total_sentences*$)$, so the coefficient can be interpreted as an approximate percentage change in report length. The estimated coefficient on $ESGUI^{(12)}$ is essentially zero ($\beta = -0.001$) and statistically insignificant. This indicates that higher ESG uncertainty is not associated with systematically shorter or longer annual reports once fund and report-year fixed effects (and the funding ratio) are controlled for.

This result reduces the concern that the main ESGUI effects operate mechanically through changes in report verbosity. Nevertheless, report length may still be correlated with other unobserved fund-year factors that also affect disclosure. For this reason, specifications that explicitly control for $\log($*total_sentences*$)$ are treated as robustness checks rather than as the preferred baseline. Table A.14 therefore re-estimates the breadth regressions while including $\log($*total_sentences*$)$ as an additional control.

As expected, report length is strongly positively associated with all three breadth measures: longer reports mention more SI topics, strategies, and asset classes. Importantly, however, the ESGUI coefficients remain very similar in sign, magnitude, and statistical significance compared to the baseline specification. The negative associations for *Spectrum*, *Variety*, and *Scope* therefore persist even after accounting for overall report verbosity. This suggests that the contraction in SI disclosure breadth under higher ESG-related uncertainty reflects selective narrowing of the topics and strategies discussed rather than a mechanical consequence of shorter reports.

Window length robustness

Table A.15 assesses whether the estimated relationship between ESG-related uncertainty and SI disclosure is sensitive to how ESGUI is timed. The baseline regressions are re-estimated using alternative constructions of ESGUI: a 6-month rolling mean, a 9-month rolling mean, the 12-month rolling mean (baseline), and a "writing-stop" scenario measured over months $t - 14$ to $t - 3$, excluding the period immediately before publication (where t denotes the publication month).

Across windows, ESGUI coefficients are generally negative for the breadth measures, but statistical precision varies. Shorter windows (6m/9m) yield weaker and less precise estimates, while the 12-

Table A.14: Breadth Measures with Verbosity Control: $\log($*total_sentences*$)$

Dependent variable:	(1) Spectrum	(2) Variety	(3) Scope
ESGUI (12m)	-0.141*	-0.067*	-0.133**
	(0.076)	(0.037)	(0.061)
$\log($*total_sentences*$)$	3.920***	1.591***	3.285***
	(1.014)	(0.554)	(0.996)
Funding Ratio	-0.037	0.106	-0.113
	(0.145)	(0.122)	(0.168)
Fund FE	Yes	Yes	Yes
Report-year FE	Yes	Yes	Yes
Observations	781	781	781
Within R^2	0.058	0.032	0.055

Notes: The table reports coefficients from OLS regressions of Spectrum, Variety, and Scope on ESGUI, Funding Ratio, and $\log($*total_sentences*$)$ as a control for report length. Fund and year fixed effects are included. Standard errors (in parentheses) are two-way clustered by fund and publication year-month. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-group variation explained by the included regressors. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

month window produces the clearest negative associations for *Spectrum*, *Variety*, and *Scope*. The writing-stop window retains a negative association for *Spectrum* but weaker evidence for *Variety* and *Scope*. Overall, these comparisons suggest that the baseline relationships are more consistent with uncertainty measured over a longer horizon aligned with annual reporting cycles rather than short-run fluctuations.

Functional-form robustness for bounded disclosure outcomes

Table A.16 checks whether the baseline H1 findings are sensitive to modelling bounded disclosure outcomes with a nonlinear fractional-response specification rather than OLS. Using fractional logit models, the estimated marginal effects of publication-aligned ESGUI remain negative for *Intensity* and the breadth measures (*Spectrum*, *Variety*, and *Scope*), with statistically significant effects for *Spectrum*, *Variety*, and *Scope* and a marginally significant effect for *Intensity*. In contrast, the effect on *Specificity* remains small and statistically insignificant. Overall, these results reinforce the main interpretation that higher ESG uncertainty is associated primarily with modest contraction in the salience and breadth of SI disclosure, rather than a systematic shift in within-SI concreteness, and suggest that this pattern is not an artefact of treating bounded outcomes as linear.

Quadratic nonlinearity test

Table A.17 reports quadratic specifications for all five H1 outcomes as a generic check for nonlinearity in the conditional-mean relationship between ESG uncertainty and SI disclosure. Across outcomes, the squared term is generally imprecisely estimated, suggesting limited evidence that the baseline relationships are strongly nonlinear. The main exception is Variety, where the negative linear term and positive squared term imply a U-shaped pattern. This evidence should be interpreted cautiously because the identifying variation in $ESGUI^{(12)}$ is primarily within-year and relatively limited, so nonlinearities are difficult to estimate precisely.

Table A.15: Timing robustness: alternative ESGUI windows

Dep. var.	(1) Intensity	(2) Spectrum	(3) Variety	(4) Scope	(5) Specificity
ESGUI 6-month rolling mean					
ESGUI (6m)	-0.001	-0.031	-0.048*	-0.103*	-0.002
	(0.001)	(0.056)	(0.028)	(0.055)	(0.002)
Funding Ratio	-0.005*	-0.099	0.085	-0.154	-0.006
	(0.003)	(0.160)	(0.122)	(0.165)	(0.006)
Within R^2	0.015	0.001	0.005	0.009	0.006
Observations	781	781	781	781	781
ESGUI 9-month rolling mean					
ESGUI (9m)	-0.001	-0.071	-0.038	-0.110**	-0.002
	(0.001)	(0.055)	(0.030)	(0.052)	(0.002)
Funding Ratio	-0.005*	-0.098	0.080	-0.162	-0.006
	(0.003)	(0.160)	(0.121)	(0.165)	(0.006)
Within R^2	0.017	0.003	0.003	0.010	0.005
Observations	781	781	781	781	781
ESGUI 12-month rolling mean (baseline)					
ESGUI (12m)	-0.002*	-0.145**	-0.069*	-0.137**	-0.003
	(0.001)	(0.071)	(0.036)	(0.062)	(0.002)
Funding Ratio	-0.005*	-0.088	0.085	-0.156	-0.006
	(0.003)	(0.160)	(0.122)	(0.165)	(0.006)
Within R^2	0.022	0.008	0.006	0.011	0.007
Observations	781	781	781	781	781
ESGUI writing-stop mean ($t-14$ to $t-3$)					
ESGUI (writing stop)	-0.002	-0.147*	-0.046	-0.119	-0.002
	(0.001)	(0.080)	(0.051)	(0.074)	(0.003)
Funding Ratio	-0.005*	-0.101	0.078	-0.168	-0.006
	(0.003)	(0.159)	(0.120)	(0.165)	(0.006)
Within R^2	0.022	0.007	0.003	0.008	0.005
Observations	781	781	781	781	781
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: Each block reports a separate regression that differs only in the ESGUI timing window. Standard errors are two-way clustered by fund and publication year-month and reported in parentheses. The writing-stop mean is computed over months $t-14$ to $t-3$ (with t denoting the publication month). Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-group variation explained by the included regressors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.16: Functional-Form Robustness: Fractional Response Models for Bounded Disclosure Outcomes

Dep. Var.	Intensity (Frac. Logit)	Spectrum/12 (Frac. Logit)	Variety/5 (Frac. Logit)	Scope/8 (Frac. Logit)	Specificity (Frac. Logit)
ESGUI (12m)	-0.001	-0.168**	-0.080**	-0.117*	-0.002
	(0.001)	(0.079)	(0.034)	(0.062)	(0.001)
Funding Ratio	-0.003*	-0.065	0.088	-0.124	-0.005
	(0.002)	(0.167)	(0.083)	(0.166)	(0.005)
Fund FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
Pseudo R^2	0.069	0.248	0.230	0.271	0.076

Notes: The table reports average marginal effects (AMEs) from fractional logit models (GLM with binomial family and logit link) following Papke et al. (1996). For Spectrum, Variety, and Scope, outcomes are rescaled to the unit interval prior to estimation (Spectrum/12, Variety/5, Scope/8) and marginal effects are rescaled back to original units for interpretability. Standard errors (in parentheses) are obtained via fund-level block bootstrap (500 replications) to account for within-fund clustering. Fund and year fixed effects are included. Pseudo R^2 is McFadden's pseudo R^2 computed as $1 - \mathcal{L}_{\text{full}}/\mathcal{L}_{\text{null}}$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.17: Test for nonlinearity: quadratic specifications

Dep. Var:	(1) Intensity	(2) Spectrum	(3) Variety	(4) Scope	(5) Specificity
ESGUI (12m mean)	-0.002	-0.296	-0.459**	0.060	-0.018
	(0.004)	(0.325)	(0.182)	(0.279)	(0.011)
ESGUI (12m mean)^2	0.000	0.002	0.004**	-0.002	0.000
	(0.000)	(0.004)	(0.002)	(0.003)	(0.000)
Funding ratio (10pp)	-0.005	-0.091	0.078	-0.153	-0.006
	(0.003)	(0.159)	(0.121)	(0.166)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Report-year FE	Yes	Yes	Yes	Yes	Yes
Observations	781	781	781	781	781
Within R^2	0.022	0.009	0.013	0.012	0.010

Notes: All models include fund and report-year fixed effects. Standard errors are two-way clustered by fund and publication year-month and reported in parentheses. Within R^2 is computed from the two-way fixed-effects transformation (fund and year) and reflects the share of within-group variation explained by the included regressors. Using the estimated coefficients, the implied turning point for Variety is approximately 52.1 ESGUI (U). Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Joint moderation test

To account for the concern that carbon exposure correlates with fund capacity (size), and that apparent WACI heterogeneity could partly reflect AUM-based differences documented in H2, I also estimate a joint moderation model that includes interactions of $ESGUI_{it}^{(12)}$ with both $\log($*AUM*$)$ and $\log($*WACI*$)$. The purpose is to assess whether the WACI-based heterogeneity reflects carbon exposure in its own right or partly proxies for size-related capacity differences. Table A.18 shows the results of this estimation with two interaction terms. The negative $ESGUI \times \log(WACI)$ interaction for *Variety* persists with nearly identical magnitude and strong statistical significance, indicating that the carbon-exposure moderation for strategic breadth is not driven by fund size.

Table A.18: Carbon Exposure (WACI) Moderation Controlling for Fund Size (AUM)

	(1)	(2)	(3)	(4)	(5)
Dep. var.	Intensity	Spectrum	Variety	Scope	Specificity
ESGUI (12m)	-0.002	-0.127*	-0.054	-0.133**	-0.002
	(0.001)	(0.075)	(0.037)	(0.065)	(0.002)
ESGUI × log(AUM)	0.000***	0.002	0.003	0.014***	0.001***
	(0.000)	(0.004)	(0.005)	(0.004)	(0.000)
ESGUI × log(WACI)	0.000	-0.020	-0.025***	-0.009	0.001**
	(0.000)	(0.013)	(0.006)	(0.013)	(0.000)
Funding Ratio	-0.003*	-0.114	0.063	-0.071	-0.005
	(0.002)	(0.171)	(0.118)	(0.158)	(0.006)
Fund FE	Yes	Yes	Yes	Yes	Yes
Report-year FE	Yes	Yes	Yes	Yes	Yes
Observations	727	727	727	727	727
Within R^2	0.069	0.011	0.023	0.030	0.035

Notes: Coefficients from OLS regressions with fund and report-year fixed effects. Standard errors are two-way clustered by fund and publication year-month and shown in parentheses. Within R^2 is the two-way fixed-effects within R^2 (fund and report-year). *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

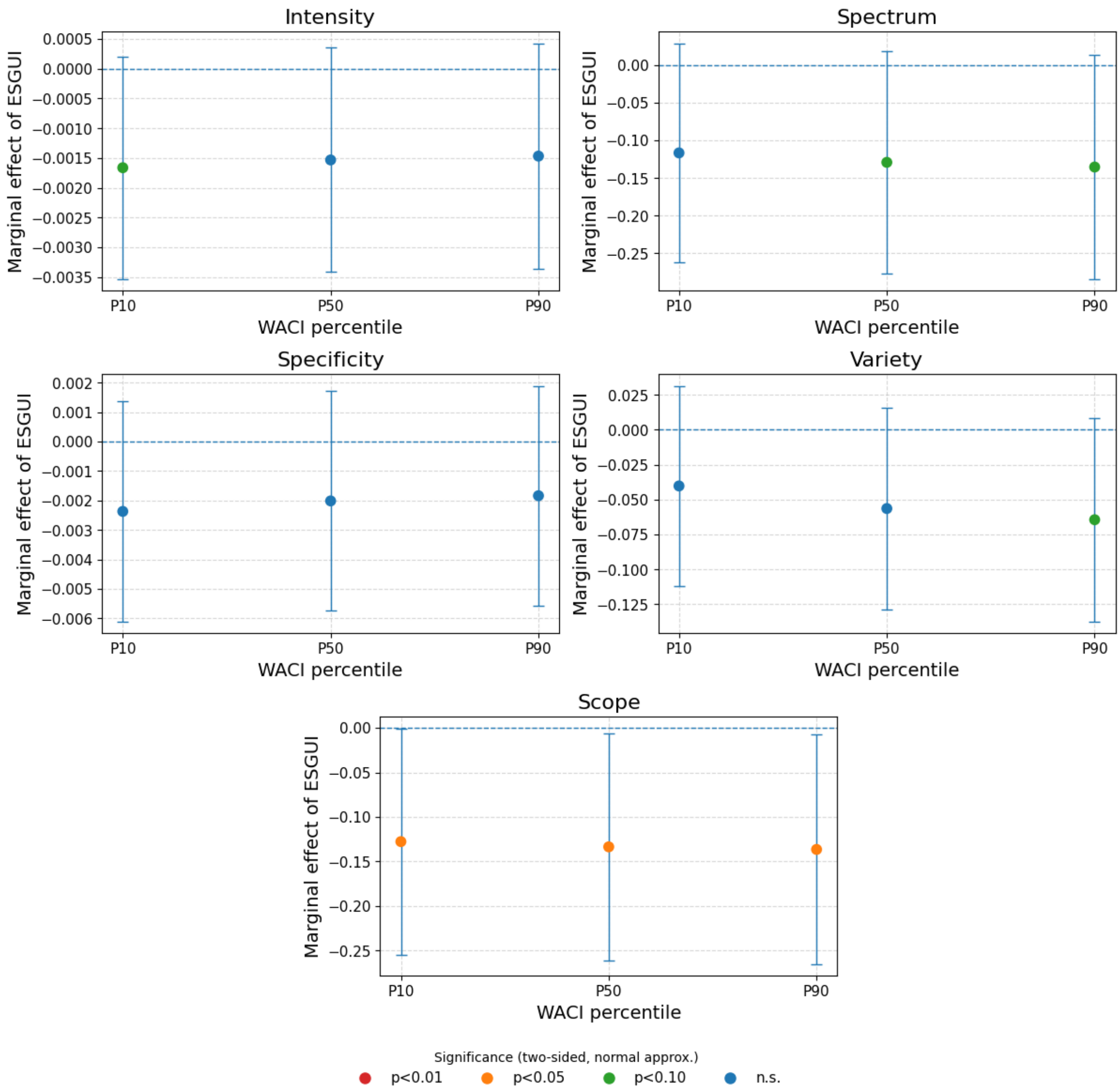

Figure A.2: Marginal effects of ESGUI on SI measures across carbon exposure (WACI) controlling for fund size (AUM). The figure reports estimated marginal effects of ESGUI (12-month rolling mean) evaluated at the 10th, 50th, and 90th percentiles of funds' carbon exposure (WACI). Points indicate marginal effects and whiskers denote 95% confidence intervals. The horizontal dashed line marks zero. Point colours reflect two-sided significance levels based on a normal approximation (p<0.01, p<0.05, p<0.10, n.s.).

A.4. Supervisory Recommendations

This appendix offers practical recommendations for financial supervision based on the findings of this thesis. These suggestions aim to help supervisors monitor SI disclosures by pension funds, especially when ESG debates intensify. Although the empirical results should be viewed with caution because of data limitations, they indicate that narrative sustainability disclosures can react to external debate and uncertainty. Supervisors may find it useful to watch how disclosure patterns change over time.

Monitor narrative disclosure trends systematically

Pension funds often share information about sustainable investing in the narrative parts of their annual reports. This thesis shows that what they disclose and focus on can shift when there is more uncertainty in the ESG debate. Supervisors might therefore benefit from tracking how these narrative sustainability disclosures change across funds and over time. Text analysis tools can help by measuring the amount of sustainability content, the types of investment strategies discussed, and which asset classes include sustainability information. These indicators would not replace traditional supervisory reviews, but they could help identify early changes in how funds report on sustainability.

Interpret changes in disclosure with caution

An important premise in this thesis is that changes in disclosure do not always reflect changes in how investments are managed. Pension funds may report less about sustainability during periods of political or reputational debate, even if their investment strategies remain unchanged. Supervisors should therefore be careful when interpreting decreases in narrative disclosure and avoid assuming this implies less sustainable investment activity. Reviewing narrative disclosures together with other sources, such as portfolio data or engagement reports, can provide a more complete view of how pension funds handle sustainability.

Provide clear expectations for sustainability reporting

When pension funds are unsure how to communicate their sustainability policies and strategies, they may become overly cautious in their reporting. Setting clear expectations for sustainability disclosure can help reduce this uncertainty. Guidance documents, thematic reviews, and sector communications can show pension funds what supervisors expect them to report and how to explain sustainability risks. While this guidance does not need to dictate exact wording, it clarifies which aspects of sustainable investing should be covered in annual reports.

Pay attention to differences between funds

The findings show that smaller pension funds tend to change their sustainability disclosures more when the ESG debate becomes more active. This could be because they have fewer resources, less reporting capacity, or greater concerns about their reputation. Supervisors may therefore monitor whether smaller funds struggle to maintain consistent sustainability reporting. Sharing examples of good reporting or offering insights from the sector could help support funds with limited resources for sustainability reporting.

Use disclosure indicators as a monitoring tool

The indicators developed in this thesis show that text analysis can capture different aspects of sustainability disclosure, including focus, scope, and specificity. Supervisors could use similar indicators to see if changes in regulation, guidance, or public debate are reflected in how pension funds report on sustainability. Over time, these indicators could help spot sector-wide reporting trends, notice unusual changes in disclosure patterns, and assess whether supervisory actions lead to more consistent sustainability reporting.

Support further research and data availability

Finally, improving data availability can strengthen academic research and supervisory monitoring. Connecting narrative disclosures with details about portfolio holdings, engagement activities, or voting behaviour makes it easier to see if changes in sustainability reporting match changes in investment practices. Promoting transparency and consistent reporting standards supports both supervision and sustainable finance research.