frontiers in Sustainability

TYPE Original Research
PUBLISHED 24 September 2026
DOI 10.3389/frsus.2026.1954577

Check for updates

OPEN ACCESS

EDITED BY
Maurício Leite,
Regional Community University of Chapecó, Brazil

REVIEWED BY
Sapna Malya,
S. P. Jain Institute of Management and Research, India
Rima Mondal,
Indian Institute of Management Calcutta, India

*CORRESPONDENCE
Mohsin Khan
✉ mohsin.khan@vit.ac.in

RECEIVED 31 July 2026
REVISED 28 August 2026
ACCEPTED 01 September 2026
PUBLISHED 24 September 2026
CORRECTED 05 October 2026

CITATION
Safas VP and Khan M (2026) Signaling ESG disclosure readiness before listing: text-mining evidence from Indian IPO prospectuses. *Front. Sustain.* 7:1954577. doi: 10.3389/frsus.2026.1954577

COPYRIGHT
© 2026 Safas and Khan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Signaling ESG disclosure readiness before listing: text-mining evidence from Indian IPO prospectuses

VP Safas and Mohsin Khan*

Vellore Institute of Technology, Vellore, Tamil Nadu, India

Introduction: Environmental, social, and governance (ESG) disclosure has become central to how firms establish credibility with capital markets, yet its role at the point of market entry remains underexamined. This study investigates how Indian firms communicate ESG readiness in mainboard initial public offering (IPO) prospectuses, measuring the intensity and thematic composition of ESG-related language rather than treating disclosure as a proxy for ESG performance or quality.

Methods: The analytical sample comprises 310 non-financial operating-company IPO prospectuses issued from FY2021-22 to FY2025-26. Prospectus text was cleaned, segmented into sentences, and embedded using Sentence-BERT. Each sentence was matched by cosine similarity to reference sentences derived from a Baier et al.-based ESG taxonomy comprising three pillars, ten categories, and forty subcategories; sentences scoring 0.45 or above were classified as ESG-related and assigned to the corresponding pillar and subcategory. In addition to the textual analysis, the study also carried out firm-level regressions relating pillar-specific disclosure to issuer characteristics and board structure, controlling for sector, year and document length.

Results: ESG-related sentences constitute, on average, 11.382% of total sentences across the sample. Governance disclosure is significantly more prevalent (8.605% of clean sentences) than social (1.462%) or environmental (1.314%) disclosure, with governance accounting for 75.5% of all ESG-classified sentences. Disclosure intensity and thematic emphasis vary meaningfully across financial years, industry sectors, and ESG subcategories. Supplementary regressions reveal pillar-specific rather than uniform associations between issuer characteristics and ESG disclosure, while board size and board independence are jointly associated with governance disclosure in a parsimonious specification.

Discussion: These findings are consistent with, but do not establish, a credibility-related signaling channel centered on governance disclosure. In the IPO prospectuses governance-related content dominates ESG communication at listing, while environmental and social disclosure remain comparatively limited and unevenly distributed across sectors and time.

KEYWORDS
corporate governance, ESG disclosure, Indian capital markets, IPO prospectus, sentence-BERT, signaling theory, text mining

1 Introduction

ESG disclosure has evolved from being a marginal activity in corporate communication into one of the essential ways in which firms can build credibility with their stakeholders, such as capital markets, regulators, and citizens (Kócziás and Demirel, 2025; Suieubayeva et al., 2025). With increasing emphasis on sustainability in investment decisions by institutional investors, firms in developed as well as emerging markets are under growing pressure to formally disclose their practices with regard to environmental, social, and governance (Friede et al., 2015; Pedersen et al., 2021). The situation is especially serious in connection with the process of changing the form of ownership of firms, in particular when they make the transition to a publicly owned status, as the IPO prospectus is the first document that will be subject to scrutiny and which the issuer uses to communicate to the uncertain investor base their business model and organizational practices (Bhabra and Pettway, 2003; Healy and Palepu, 2001). Unlike the annual reports and sustainability reports of already-listed firms, which have been extensively studied using text-mining and disclosure-scoring methods (Baier et al., 2020; Heichl and Hirsch, 2023; Menon et al., 2026) IPO prospectuses have received comparatively little attention as ESG-bearing documents, despite being issued precisely at the moment when information asymmetry between the firm and outside investors is at its peak (Derrien, 2005; Ritter and Welch, 2002).

The theoretical rationale for expecting ESG content in IPO prospectuses rests on well-established strands of the finance and management literature. Signaling theory holds that firms possessing genuine underlying quality have an incentive to transmit costly, difficult-to-imitate signals that separate them from lower-quality issuers (Connelly et al., 2011). In the IPO context, voluntary or elaborated ESG disclosure functions as one such signal, allowing issuers to communicate governance readiness, regulatory preparedness, and long-term orientation to prospective shareholders (Allen and Faulhaber, 1989; Benveniste and Spindt, 1989; Del Gesso and Lodhi, 2025). To this end, legitimacy theory holds that firms provide information about their environmental and social behavior as a means of demonstrating compliance with the prevailing social norms, thereby obtaining a social license to operate, especially in areas where there is increasing pressure for regulation (Guthrie et al., 2007; Suchman, 1995) In addition, stakeholder theory predicts that firms react not only to the needs of the investors, but also those of a larger number of regulators, employees, communities, and civil society groups whose expectations influence disclosure decisions (Gallego-Álvarez and Ortas, 2017; Nirino et al., 2021). Institutional theory provides an additional perspective on the subject in the sense that, as ESG disclosure becomes institutionally and coercively embedded in a regulatory field, firms tend to exhibit similar behaviors in regard to disclosure despite differences in their sustainability. Taken together, these frameworks imply that IPO prospectuses are not incidental repositories of ESG language but strategically constructed market-entry documents whose content merits systematic empirical investigation.

Emerging evidence, largely drawn from developed markets, indicates that ESG-related disclosure in offering documents is associated with meaningful capital-market outcomes. Fu et al. (2023) find that United States IPO firms with superior ESG disclosure exhibit a higher probability of post-listing survival, while Kang and Lam (2023) show that environmental disclosure is linked to reduced IPO underpricing in Singapore, and Reber et al. (2022) document a relationship between ESG disclosure and idiosyncratic risk among newly listed firms. In the Indian setting specifically, Kirubakaran and Joseph (2026) report a significant positive association between an aggregate, dictionary-based ESG score constructed from IPO prospectuses and listing-day as well as short-horizon post-listing returns, a finding broadly consistent with the signaling logic outlined above. Yet this literature, including the aggregate-score approach, treats ESG disclosure largely as a unidimensional construct; more disclosure is treated as more signal, without systematically decomposing that signal into its constituent environmental, social, and governance pillars, or the finer thematic subcategories within each pillar, and their approach also counts the keywords as ESG disclosure. This is a non-trivial omission. Evidence from listed-firm sustainability reporting consistently shows that the environmental, social, and governance dimensions of disclosure are not interchangeable: they respond to different regulatory drivers, are emphasized differently across sectors, and carry different associations with financial and non-financial outcomes (Alodat and Hao, 2025; Bernard et al., 2025; Shobhwani and Lodha, 2024). A single aggregate ESG percentage computed from a prospectus may therefore obscure whether an issuer is signaling genuine environmental stewardship and social responsibility, or is instead concentrating its disclosure narrative around governance and compliance content that is comparatively easy to verify and, in many jurisdictions, partially mandated by company law and securities regulation (Bae et al., 2018; Hummel and Schlick, 2016).

The problem is especially pronounced in India because although the regulatory environment with respect to ESG disclosures in India has been developing very quickly, its focus has always been on listed firms and not initial public offering firms. The Securities and Exchange Board of India developed the Business Responsibility and Sustainability Reporting (BRSR) framework in 2021, improved it in 2023 by introducing BRSR Core Assurance and Value Chain Disclosure Requirements, and has continuously worked toward improving the framework by making several more changes (Securities and Exchange Board of India, 2021, 2023, 2025). Because the requirements apply after listing, ESG disclosures in the IPO prospectus are voluntary, thus providing a convenient laboratory in which to study how issuers choose to frame their ESG position in the absence of a standard mandatory disclosure format (Sinha Ray and Goel, 2023; Veeravel et al., 2024). On the other hand, findings from other transition regulatory contexts indicate that corporate ESG communication is usually uneven, with language around governance and compliance generally becoming solidified before the emergence of environmental and social language, while there is noticeable variation between industries based on materiality and regulatory exposure (Lee et al., 2026). It has yet to be systematically explored whether such a pattern applies to the ESG communication in the prospectuses of IPOs in India—documents drafted amid intense legal review and due diligence on the part of the underwriters, as well as read by the first generation of public investors.

In terms of methodology, previous research analyzing ESG language in Indian corporate disclosure has focused mainly on third-party ESG ratings, which are secretive, varied between rating agencies, and not available for newly listed companies (Heichl and Hirsch, 2023; Bernard et al., 2025), or on keyword-based dictionary counts, which cannot distinguish between genuine thematic relevance and superficial lexical overlap and are consequently vulnerable to both false positives and false negatives (Fenili and Raimondo, 2021; Loughran and McDonald, 2013). Recent advances in natural language processing offer a more precise alternative. Sentence-BERT and related

transformer-based sentence-embedding models capture the semantic meaning of text rather than relying on exact keyword matches, enabling the classification of prospectus sentences into ESG themes even where the surface vocabulary departs from a fixed dictionary. Such semantic approaches have recently been applied to annual report disclosures in the Indian hospitality sector (Menon et al., 2026) and to large-scale corpora of Korean sustainability reports (Lee et al., 2026), demonstrating that sentence-level embedding classification can recover interpretable, temporally and sector differentiated patterns of ESG emphasis that dictionary counts and coarse aggregate scores tend to mask. To date, however, this class of method has not been applied to IPO prospectuses, and no study has combined a validated, multi-pillar ESG taxonomy with sentence-level semantic classification to characterize the internal composition rather than merely the volume of ESG disclosure at the point of market entry.

Against this backdrop, the present study addresses a threefold gap in the literature. First, it extends the study of ESG disclosure beyond the post-listing annual report and sustainability report literature (Baier et al., 2020; Menon et al., 2026; Shobhwani and Lodha, 2024) to the IPO prospectus, a market-entry document prepared under distinctive legal, temporal, and information-asymmetry conditions (Crain et al., 2021; Oware and David Kweku Botchway, 2023). Second, it moves beyond treating ESG disclosure as a single aggregate percentage, as in prior Indian IPO research (Kirubakaran and Joseph, 2026), by decomposing disclosure into its environmental, social, and governance pillars and into forty underlying subcategories derived from an established taxonomy (Baier et al., 2020), thereby enabling an assessment of which themes—rather than simply how much issuers choose to emphasize. Finally, it contributes the use of a semantic sentence-BERT classifier to the field of IPOs, which is a more consistent and reproducible way of measuring disclosure intensity as compared to dictionary measures, but still scalable for large datasets (Reimers and Gurevych, 2019). With its measurement of the disclosure intensity and composition of pillars over a five-year span of IPOs from the Indian main board, during the period of regulatory change, the paper also addresses the larger issue of whether voluntary communication of ESG issues in the offer documents takes place in anticipation of or along with post-listing disclosure obligations (Grewal et al., 2019; Zhang et al., 2023).

Implications of the findings extend to various groups of stakeholders. To securities regulators and stock exchanges, knowledge of the present make-up of ESG language in prospectuses will help in deciding if the disclosure guidelines should oblige the issuers to make distinctions between the environmental, social and governance aspects instead of allowing a single and undifferentiated ESG narrative where the boilerplate disclosures in governance aspects are used as a substitute for any environmental or social commitment (Manes-Rossi and Nicolo', 2022; Monteiro et al., 2023). To institutional investors, underwriters and analysts, the pillar-based metric provides a more meaningful approach to judge an issuer's sustainability stance than a summary score, as evidenced by the differences between disclosure quantity and quality of disclosure (Hummel and Schlick, 2016; Michelon et al., 2015). For researchers, the research offers a reproducible and taxonomy-based methodology to measure ESG disclosure in the offering document, which can be further applied in other developing countries going through similar regulatory changes (Debnath et al., 2025; Kamugisha and Sun, 2025).

As such, the paper focuses on the following research questions: RQ1 explores the degree and pillar structure of ESG disclosures in IPO prospectuses of Indian issuers; RQ2 analyses the variation of such disclosures over financial years and industry sectors; and RQ3 discovers the dominating ESG sub-categories in the prospectus disclosures and their variation in terms of time and sectorial differences. RQ4 examines whether observable issuer characteristics and board structure are associated with the intensity and composition of environmental, social, and governance disclosure in IPO prospectuses. The rest of the paper is organized as follows. In Section 2, we develop the theoretical framework and review the relevant literature on ESG disclosure, signaling in the IPO market, and text mining sustainability reporting. Section 3 provides the description of the sample of Indian main board IPOs and the sample selection process. Section 4 explains our text extraction process, the taxonomy, the Sentence-BERT classification approach, and the scoring method. Section 5 provides our results, and Section 6 discusses the results in connection with signaling, legitimacy, and institutional theories, followed by sections 7& 8 discussing implications/limitations and conclusion.

2 Literature review and theoretical framing

2.1 Theoretical framework

There are three theoretical approaches that together support the expectation that ESG-related language will be systematically included in the IPO prospectus: signalling theory, legitimacy theory, and institutional theory. First, signalling theory is based on the fact that in situations involving information asymmetry, there is an incentive for the better-informed individual to send costly and non-replicable signals of quality (Spence, 1978). In the context of capital markets, it has been shown by Connelly et al. (2011) that signalling theory has become the predominant approach for analysing the motives behind voluntary disclosure because of the difficulty of fabricating such information by companies with worse prospects. In the context of IPOs, this has led to the development of the deliberate-underpricing signal by Allen and Faulhaber (1989) and the price discovery process through bookbuilding by Benveniste and Spindt (1989). Firms with favorable prospects try to differentiate themselves by means of costly and verifiable signals. Finally, in a systematic survey of literature related to the theory of ESG disclosure, Del Gesso and Lodhi (2025) find that signalling theory is the most common motivation cited by researchers for voluntary ESG disclosure, due to its difficulty for lower-quality firms to replicate.

The legitimacy perspective provides an equally complementary but outward-looking explanation. According to Suchman (1995) Legitimacy refers to "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, and beliefs," while Guthrie et al. (2007) have applied this to the context of environmental and social reporting, to the effect that companies provide this information to bridge the gap between social expectations and corporate behavior. In their recent study, Sehgal et al. (2022) use this idea to demonstrate that the provision of environmental and social information can act as a legitimate tool for the company operating under increased regulatory and stakeholder scrutiny in the Indian context. For a new-listing company, the IPO prospectus must be considered as the single most important legitimacy-building document because it is

through this document that the company makes its first public introduction to its shareholders, regulators, and the financial press. The institutional perspective, drawing on the work of DiMaggio and Powell (1983) this adds another level of structure to this explanation in that, as pressures increase in a field around ESG reporting, firms tend to converge into a common pattern of disclosure. This institutional lens is useful for interpreting cross-sectional similarity in prospectus disclosure that may otherwise be mistaken for evidence of genuine convergence in sustainability conduct.

Finally, stakeholder theory fills this picture by widening the audience for the disclosure of ESG information. In their article, Nirino et al. (2021) prove that ESG practices mediate the effect of corporate controversies on financial performance, thus supporting the claim that corporations have to disclose certain information in order to meet the expectations of diverse groups of stakeholders, including employees, regulators, communities, and civil society organisations, not just shareholders. Gallego-Álvarez and Ortas (2017) prove that both stakeholder pressure and cultural context affect the intensity of corporate social disclosure in different countries, and Nicolò et al. (2022) demonstrate that such factors as the demographic composition of the board influence the extent of ESG disclosure among European corporations. Thus, combining these three approaches, one can conclude that the presence of ESG information in the IPO prospectus does not represent a neutral result of a long document; instead, it reflects the decision made by corporations based on the need to communicate the quality of their products to investors and legitimise themselves before other interested parties.

2.2 ESG disclosure, signalling, and IPO market outcomes

However, a recent stream of studies examines the relationship between ESG disclosure and IPO prices and performances, even though such studies are still primarily undertaken in developed countries. Specifically, Fu et al. (2023) show that ESG disclosure leads to an increased chance of survival of US IPO firms after going public, thus supporting the hypothesis that ESG disclosure acts as a costly signal of organisational robustness. In turn, Kang and Lam (2023) have discovered that there is a negative association between environmental disclosure intensity and the level of IPO underpricing in Singaporean companies. Additionally, Reber et al. (2022) reveal that ESG disclosure intensity is correlated with idiosyncratic risks of IPOs. Within India specifically, Kirubakaran and Joseph (2026) Construct a dictionary-based ESG score from the prospectuses of 344 mainboard IPOs and find a significant positive relationship between the aggregate ESG score and both listing-day returns and short-horizon post-listing performance, with the effect persisting up to fifteen days before attenuating—an outcome the authors interpret as evidence that ESG disclosure operates as a short-term signal whose informational value is gradually supplanted by financial performance data and market-based ESG ratings once a firm is publicly traded.

These IPO-specific findings are part of a wider literature relating firm-level ESG disclosures to underpricing and prospectus contents. Crain et al. (2021) reveal how the contents of a prospectus can affect IPO pricing by alleviating information asymmetry problems in line with classical determinants of underpricing highlighted by Beatty and Ritter (1986), Carter and Manaster (1990), Carter et al. (1998); and the market sentiments and oversubscription effect found for IPOs in India by Sahoo and Rajib (2010), Bhattacharya et al. (2020), Babu and Dsouza (2021). Veeravel et al. (2024) examine the relationship between ESG and the financial performance of Indian listed companies and observe support for investors, increasing appreciation of disclosure quality, whereas Abbas et al. (2025) find that community and environment disclosure (and not ESG disclosures as a whole) affects the long-run performance of stocks issued by IPOs in Malaysia. This pillar-specific result is significant because it suggests that the pieces of information disclosed about E, S, and G do not necessarily carry equal value, an empirical question that cannot be answered using the aggregation approach (dictionary counting included).

An additional stream of research warns against viewing the extent of disclosure as synonymous with its quality and the firm's actual ESG performance. Michelon et al. (2015) demonstrate that the adoption of typical practices of CSR reporting by firms does not imply the provision of balanced or meaningful disclosure, while the findings emerging from research into signalling-based governance (Bae et al., 2018) point out that voluntary sustainability disclosure by South Asian firms is driven by their governance structure to the same extent as their actual sustainability practices. Moreover, Arena et al. (2015) argue that the informativeness of environmental disclosure tone depends on good governance quality. All of the above-mentioned studies imply that it is necessary to examine the content of an aggregate ESG disclosure percentage, especially within a prospectus framework, where the proportion of governance-related content (board structure, audits, related-party transactions) can be expected to be larger than that of voluntary environmental or social narrative due to statutory and merchant-banker due diligence obligations (Bhabra and Pettway, 2003; Healy and Palepu, 2001).

2.3 Text-mining and semantic approaches to ESG disclosure measurement

Methodologically, the measurement of ESG disclosure has evolved through three broad generations. The earliest approach relies on third-party ESG ratings from providers such as MSCI or Sustainalytics; however, these ratings have been criticized for limited transparency and low inter-agency agreement, and are in any case unavailable for firms without an established public trading history, making them unsuitable for newly listed or pre-IPO firms (Heichl and Hirsch, 2023; Bernard et al., 2025). A second generation relies on dictionary-based textual analysis, in which the frequency of pre-classified positive, negative, or thematic terms is counted and scaled by document length (Loughran and McDonald, 2013; Tetlock et al., 2008). This approach has been applied productively to IPO documents in the United States by Ferris et al. (2013) and Krinitz and Neumann (2021), and to ESG-specific sentiment in IPO prospectuses by Fenili and Raimondo (2021), whose dictionary forms part of the basis for ESG word lists used in subsequent Indian IPO research (Kirubakaran and Joseph, 2026). Dictionary methods are computationally efficient and transparent, drawing on the broader tradition of financial sentiment analysis established by Antweiler and Frank (2004), Cornelli et al. (2006), Seng et al. (2017), and reviewed methodologically by Shmueli and Koppius (2011) and Chhatwani (2021). However, their fundamental limitation lies in the fact that it is not possible to determine if statements are semantically equivalent, even though the keywords used are different, as well as determining if the term has any ESG significance at all.

The third generation tackles this drawback by utilising transformer-based sentence embeddings. Specifically, Reimers and Gurevych (2019) propose Sentence-BERT (SBERT), which generates sentence embeddings that have a high level of semantic meaning and a constant size and that are thus easy to compare using cosine similarity. Building on the structured ESG concept taxonomy developed by Baier et al. (2020), which organises ESG-related vocabulary into pillars, categories, and subcategories on the basis of the United States 10-K and proxy-statement disclosure, recent studies have used SBERT-based semantic matching to classify corporate disclosure text against reference concepts rather than fixed keyword lists. Menon et al. (2026) apply this approach to the annual reports of Indian hospitality and tourism firms, generating firm-year pillar-level ESG scores that are subsequently linked to accounting-based financial performance. Schimanski et al. (2024) similarly, demonstrate that natural-language-processing methods, including large language models, can quantify environmental, social, and governance communication with greater validity and scalability than dictionary counts, while cautioning that such measures capture disclosure intensity and thematic emphasis rather than verified organisational practice. At the sustainability-report level, Lee et al. (2026) apply structural topic modeling to more than six hundred Korean corporate sustainability reports and find a phased evolution in ESG discourse from broad corporate social responsibility narratives toward more technical, compliance-oriented governance and climate-risk content—together with pronounced sector-specific patterns of thematic emphasis, a finding consistent with the broader argument that sustainability disclosure should be read as a structured communicative artefact shaped by legitimacy-seeking and institutional pressures rather than as a direct proxy for firm performance (Suchman, 1995; Grewal and Serafeim, 2020; Koseoglu et al., 2021).

2.4 Synthesis and positioning of the present study

Read together, this literature establishes three propositions that motivate the present study's design. First, signalling, legitimacy, and stakeholder theory jointly predict that IPO prospectuses should contain systematic ESG content, and prior Indian evidence confirms that aggregate ESG disclosure in IPO prospectuses is priced by the market (Kirubakaran and Joseph, 2026). Second, evidence from both IPO and post-listing disclosure research indicates that the environmental, social, and governance pillars are not informationally interchangeable, and that disclosure volume can diverge from disclosure quality or substantive ESG performance (Bae et al., 2018; Michelon et al., 2015; Abbas et al., 2025). Third, semantic sentence-embedding methods anchored in an established ESG taxonomy offer a more valid and scalable alternative to both unavailable third-party ratings and coarse dictionary counts, and have already been validated on Indian corporate disclosure and cross-national sustainability-report corpora (Baier et al., 2020; Menon et al., 2026; Lee et al., 2026). No prior study, to our knowledge, has combined a multi-pillar, subcategory-level ESG taxonomy with sentence-level SBERT classification to characterize the internal composition of ESG disclosure specifically within IPO prospectuses. The present study is positioned to close this gap by decomposing prospectus-level ESG disclosure into pillar and subcategory shares across a five-year corpus of Indian mainboard IPOs, thereby testing whether the aggregate ESG signal identified in prior research is uniformly distributed across environmental, social, and governance content or is instead concentrated in a narrower set of themes.

3 Data and sample

The study covers Indian mainboard IPOs issued during the five financial years from FY2021-22 to FY2025-26 (1 April 2021 to 31 March 2026). The study period from FY2021–22 to FY2025–26 was selected to capture the contemporary post-2020 ESG regulatory transformation, during which major policy and mandatory disclosure developments reshaped corporate reporting practices (Kim and Yang, 2026), extending the analysis to earlier years could dilute or obscure the influence of these recent institutional changes on IPO prospectus disclosure. The unit of analysis is the IPO prospectus. Prospectuses are treated as market-entry documents because they are prepared at the transition from private to public ownership and contain the issuer's formal presentation of governance arrangements, material risks, regulation, stakeholder matters, and business strategy (Bhabra and Pettway, 2003).

REITs and InvITs are excluded because their trust-based asset-holding structures, business models, and disclosure formats differ from those of operating companies. Financial services firms are excluded because they are subject to sector-specific prudential, capital adequacy, risk, and regulatory disclosure requirements that reduce comparability with non-financial operating companies. Table 1 explains the sample selection criteria, and the final sample size for the current study is 310 companies. The selected companies are classified based on the NSE major sector classification.

TABLE 1 Sample selection procedure.

Selection step	Excluded	Remaining	Reason
Indian mainboard IPOs, FY2021-22 to FY2025-26	—	365	Initial population
Less: REITs and InvITs	9	356	Trust-based structures and different disclosure formats
Less: financial services firms	46	310	Sector-specific prudential and regulatory disclosure regime
Final analytical sample	—	310	Non-financial operating-company IPO prospectuses

3.1 Supplementary firm-characteristic and board data

To fulfil RQ4, additional regression analysis was carried out using the following firm-specific variables. The primary data for these financial and governance variables are manually collected and cross-verified using the CMIE Prowess database and the National Stock Exchange (NSE) website. Specifically, we control for several firm characteristics, including Firm Age (years since incorporation), Total Assets (a proxy for firm size), and Issue Size. To account for firm performance and financial health, we include Return on Assets (ROA), Return on Equity (ROE), and the Debt-to-Equity (D/E) Ratio as measures of financial leverage. Furthermore, we capture corporate governance dynamics through two key board-level metrics: Board Size (the total number of directors) and Board Independence (the proportion of independent directors on the board). Finally, to account for variations in disclosure verbosity, we include Sentence Count as a document-length control.

4 Methodology

4.1 Text extraction, cleaning, and sentence segmentation

Each prospectus PDF was converted into machine-readable text. Raw text extraction from the source PDF documents was handled using the PyMuPDF (fitz) library. To ensure high-quality inputs for the embedding model, the raw text was passed through a rigorous cleaning and filtering pipeline. The extracted text was cleaned to remove non-substantive noise, normalize formatting, and prepare the document for sentence-level analysis. The current study has opted out of the removal of stop words, as when it comes to sentence BERT, removing stop words will make the sentences hard to understand, which can affect the vector matching. The cleaned text was then segmented into sentences, producing total_clean_sentences for each prospectus. This count forms the denominator for all disclosure intensity measures. Because the score is normalized by the number of clean sentences, it measures the share of the full prospectus devoted to ESG-related content rather than the raw volume of ESG language.

4.2 ESG taxonomy and reference sentences

The classification framework is based on the three-level ESG taxonomy developed by Baier et al. (2020). The taxonomy contains three pillars, ten categories, and forty subcategories. These categories were composed of keywords, but for this study, they were changed to sentences based on the category name and the keywords available under that category. The category details are shown in Table 2, and the full list of categories, subcategories, modifications, and reference sentences is provided in Appendix A. to ensure replicability. Two labels were adapted to fit the present research context: 'UNGC Compliance' was replaced with 'Statutory governance', and 'ILO core conventions' was replaced with 'Core labor standards'. Reference sentences representing each subcategory were created and embedded using the same semantic model as the prospectus sentences.

4.3 Sentence-BERT classification

Sentence-BERT (SBERT) produces semantically meaningful sentence embeddings that can be compared efficiently using cosine similarity (Reimers and Gurevych, 2019). The analysis utilized the all-MiniLM-L6-v2 Sentence-BERT (SBERT) checkpoint to generate semantic representations of the text. By default, this specific lightweight model architecture maps sentences into a 384-dimensional dense vector space and relies on mean pooling to aggregate token-level data into a single sentence embedding. During the encoding execution, the resulting sentence embeddings were explicitly L2-normalized (normalize_embeddings = True) to facilitate accurate cosine similarity calculations downstream. Each prospectus sentence

TABLE 2 ESG taxonomy used for sentence-level classification.

Main category (pillar)	Category	Subcategories
Governance	Corporate governance	Audit and control; Board structure; Remuneration; Shareholder rights; Transparency; Talent
Governance	Business ethics	Bribery and corruption; Political influence; Responsible marketing; Whistle-blowing system
Governance	Sustainability management and reporting	Disclosure and reporting; Governance of sustainability issues; Stakeholder engagement; Statutory Governance
Environmental	Climate change	Biofuels; Climate change strategy; Emissions management and reporting
Environmental	Ecosystem Service	Access to land; Biodiversity management; Water
Environmental	Environmental management	Environmental standards; Pollution control; Product opportunities; Supply chain environmental standards; Waste and recycling
Social	Public health	Access to medicine; HIV and AIDS; Nutrition; Product safety
Social	Human rights	Community relations; Privacy and free expression; Security; Weak governance zones
Social	Labor standards	Diversity; Health and safety; Core Labor Standards; Supply chain labor standards
Social	Society	Charity; Education; Employment

The taxonomy comprises 3 pillars, 10 categories, and 40 subcategories and is adapted from Baier et al. (2020). 'Statutory Governance' replaces 'UNGC Compliance', and 'Core Labor Standards' replaces 'ILO core conventions'.

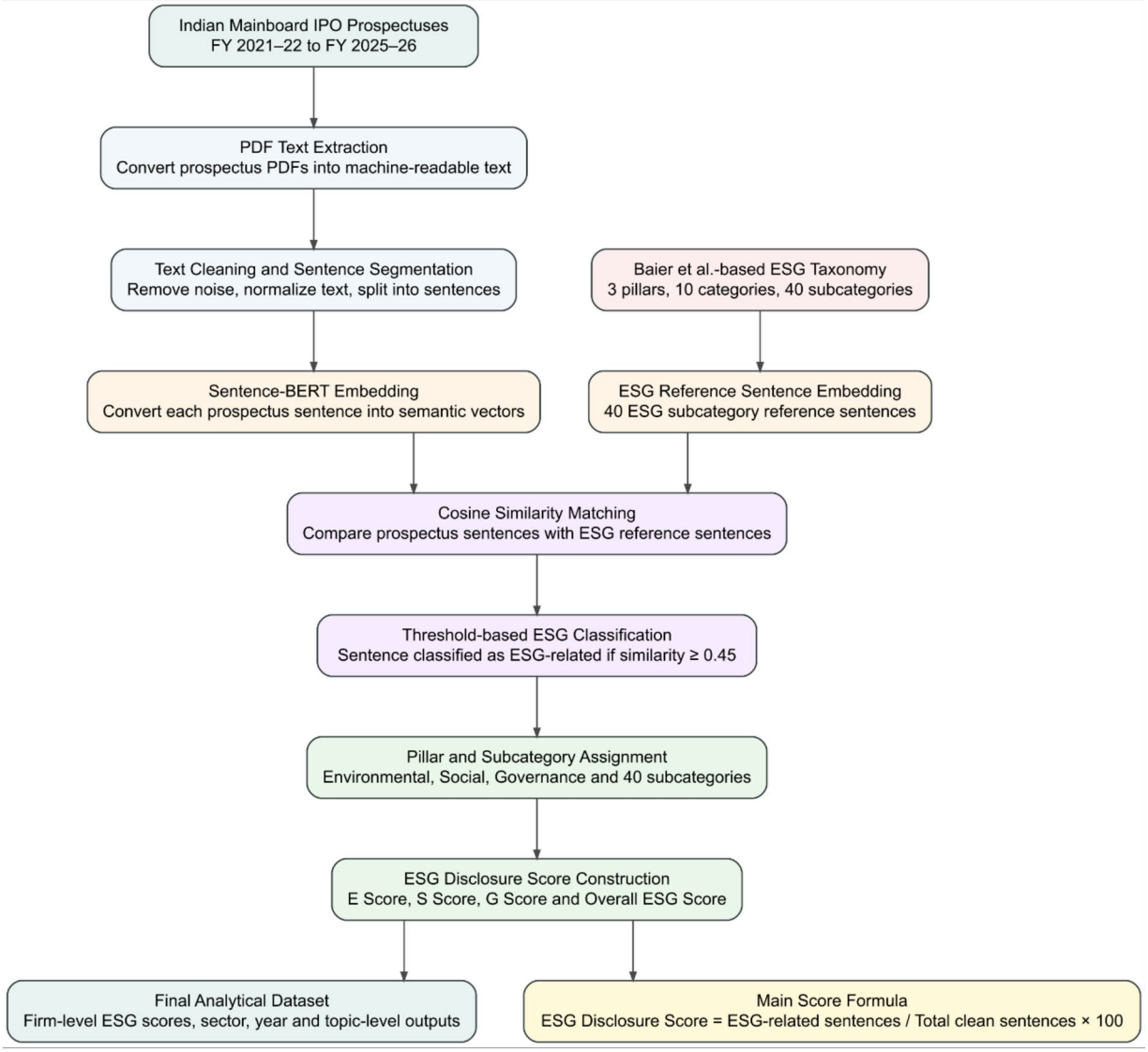


FIGURE 1
ESG score construction framework. Prospectus sentences are classified as ESG-related when their maximum cosine similarity with the ESG reference set is at least 0.45.

was converted into an embedding and compared with the ESG reference-sentence embeddings. The highest cosine similarity score determined the best-matching ESG subcategory. A sentence was classified as ESG-related when its maximum similarity score was at least 0.45. Sentences below the threshold were treated as non-ESG. Each accepted sentence was assigned to one pillar and one subcategory through the best-match rule. The detailed ESG score construction framework is shown in Figure 1.

4.4 Score construction

Disclosure intensity is calculated at the prospectus level. The overall ESG disclosure score measures the percentage of all clean prospectus sentences classified as ESG-related. Pillar scores use the same denominator and therefore indicate how much of the full prospectus is devoted to environmental, social, or governance content.

$$Overall\ ESG\ Disclosure\ Score_i = \left(ESG_count_i / total_clean_sentences_i\right) \times 100$$

$$Environmental\ Disclosure\ Score_i = \left(E_count_i / total_clean_sentences_i\right) \times 100$$

$$Social\ Disclosure\ Score_i = \left(S_count_i / total_clean_sentences_i\right) \times 100$$

$$Governance\ Disclosure\ Score_i = \left(G_count_i / total_clean_sentences_i\right) \times 100$$

Pillar composition is calculated by aggregating sentence counts across the relevant sample and dividing each pillar count by the total number of ESG-classified sentences. These measures answer a different question from the pillar disclosure scores: they show how ESG

disclosure is internally distributed after non-ESG prospectus content has been removed.

$$Environmental\ share = \left[\sum E_count / \sum ESG_count\right] \times 100$$

$$Social\ share = \left[\sum S_count / \sum ESG_count\right] \times 100$$

$$Governance\ share = \left[\sum G_count / \sum ESG_count\right] \times 100$$

Topic composition is also calculated within ESG disclosure. At the firm level, topic share equals the number of sentences assigned to a subcategory divided by the firm's ESG_count. For sector-year analysis, subcategory counts and ESG counts are aggregated within each sector-year cell before the share is calculated.

$$Topic\ share\ within\ ESG\ disclosure_{ik} = \left(subcategory_count_{ik} / ESG_count_i\right) \times 100$$

$$Sector\text{-}year\ topic\ share_{stk} = \left[\sum subcategory_count_{stk} / \sum ESG_count_{st}\right] \times 100$$

Accordingly, ESG disclosure intensity measures how much of the full prospectus is ESG-related, whereas mean topic share within ESG disclosure measures which ESG themes dominate within the classified ESG text. The latter should not be labeled 'expected topic proportions', because the method is supervised semantic matching rather than structural topic modeling.

4.5 Threshold justification, validation, and sensitivity analysis

The operational similarity threshold is 0.45. A threshold should not be justified only by convention, because classification performance depends on the embedding model, reference-sentence design, prospectus language, and class balance. For finalizing the threshold limit, the authors manually processed fifteen samples and applied different threshold limits of 0.40, 0.45, and 0.50. Afterwards, they found that during all thresholds, the number of clean sentences remained the same, even though 0.45 was able to catch more accurate sentences with fewer related sentences in the 0.40 threshold.

4.6 Descriptive, sector-year, diagnostic, and regression analysis

The analysis is conducted in four stages. Initially, prospectus-level descriptive statistics summarize document length, ESG sentence counts, and disclosure scores. Afterwards, aggregate and year-specific pillar composition is examined. Third, mean scores, distributions, and subcategory shares are compared across sectors and financial years. And finally, two diagnostic analyses are used: a scatter plot tests whether the normalized ESG score is mechanically related to prospectus length, and an ordinary least squares model estimates sector and year differences in the overall ESG disclosure score.

$$\text{ESGScore}_i = \alpha + \sum_s \beta_s \text{Sector}_{is} + \sum_t \gamma_t \text{Year}_{it} + \varepsilon_i$$

The coefficient plot uses Consumer Discretionary as the reference sector and FY2021-22 as the reference year. Coefficients are interpreted as conditional differences in disclosure intensity, not causal effects.

4.7 Supplementary analysis: issuer characteristics, board structure and ESG disclosure composition

To examine whether ESG communication varies systematically with observable issuer characteristics and board structure, we conduct a supplementary prospectus-level analysis using four dependent variables: G score, E score, S score, and firm-level G share. Issue Size, Total Assets, and Sentence Count are natural-log transformed. ROA, ROE, and D/E Ratio are winsorized at the 1st and 99th percentiles, whereas Firm Age, Board Size, and Board Independence are retained at levels. Board Independence is measured as the proportion of independent directors to total board size, and ln (Sentence Count) controls for prospectus length. Before estimation, pairwise correlations and variance inflation factors (VIFs) are examined; all substantive predictor VIFs are below 5 (maximum = 3.596).

$$\begin{aligned} Y_i = {} & \beta_0 + \beta_1 ln(Issue\ Size_i) + \beta_2 Firm\ Age_i \\ & + \beta_3 ln(Total\ Assets_i) + \beta_4 ROA_i + \beta_5 ROE_i \\ & + \beta_6 D/E\ Ratio_i + \beta_7 Board\ Size_i \\ & + \beta_8 Board\ Independence_i + \beta_9 ln(Sentence\ Count_i) \\ & + Sector\ FE + Year\ FE + \varepsilon_i \end{aligned}$$

The four equations are estimated in parallel with an identical set of predictors and fixed effects so that differences across ESG pillars can be assessed directly. For this firm-level analysis, G share is defined as G count divided by ESG count and is expressed as a proportion between 0 and 1. Accordingly, the pillar-score models capture disclosure intensity relative to the full prospectus, whereas G share captures the internal governance composition of ESG-classified text.

$$\text{G share}_i = \text{G count}_i / \text{ESG count}_i$$

To examine the credibility-related governance channel more directly, a parsimonious specification is estimated with G score as the dependent variable and Board Size and Board Independence as the focal predictors, while retaining sector and year fixed effects. Board structure is treated as an observable governance-based proxy relevant to signal credibility rather than as a direct measure of credibility itself.

$$\text{G score}_i = \beta_0 + \beta_1 \text{Board Size}_i + \beta_2 \text{Board Independence}_i + \text{Sector FE} + \text{Year FE} + \varepsilon_i$$

The Breusch-Pagan test is applied separately to each specification. When the null of homoskedasticity is rejected at the 5% level, inference is based on HC3 heteroskedasticity-robust standard errors; otherwise, conventional OLS standard errors are reported. Model reporting includes coefficient estimates, standard errors, confidence intervals, sample size, R-squared, adjusted R-squared, F-statistics, and multicollinearity diagnostics. All supplementary regressions are interpreted as conditional associations and are not intended to identify causal effects of issuer characteristics or board structure on ESG disclosure, nor to treat disclosure intensity as evidence of underlying ESG performance or firm quality.

5 Results

5.1 Descriptive statistics

An average prospectus contains around 6,365 clean sentences, of which approximately 720 are related to ESG reference sentences. The mean overall ESG disclosure score is 11.382%, indicating that around one out of ten of the prospectuses' content talks about ESG, out of which Governance disclosure accounts for 8.605% of the full prospectus on average, showcasing the main disclosed content in the ESG category, compared with 1.462% for social disclosure and 1.314% for environmental disclosure. The mean and median overall scores are close, although the range from 7.860 to 17.772% shows meaningful heterogeneity across issuers (Table 3).

Table 4 presents the descriptive statistics for the supplementary firm and board characteristics across our sample of 310 observations. On average, the firms in our sample have a board size of approximately eight directors (mean = 7.56), with independent directors making up nearly half of the board (mean = 0.49). The sampled companies are relatively established, with an average firm age of nearly 21 years, though the sample captures a wide distribution ranging from 1 to 68 years. Financially, the firms exhibit average profitability metrics of 8.6% for Return on Assets (ROA) and 17.7% for Return on Equity (ROE), alongside a moderate mean Debt-to-Equity ratio of 0.65. Finally, to mitigate the effect of extreme values and right-skewness, variables representing vast numerical ranges, specifically Issue Size, Total Assets, and our document-length control, Sentence Count, are included as natural logarithms.

5.2 Correlations and multicollinearity diagnostics

The correlation matrix (Figure 2) indicates relatively strong associations between ROA and ROE (r = 0.784) and between ln(Issue Size) and ln(Total Assets) (r = 0.772). However, the variance inflation factors (Table 5) remain below the conventional threshold of 5, with a maximum VIF of 3.596. Thus, although some predictors are moderately to strongly correlated, the diagnostics do not indicate problematic multicollinearity, and all predictors are retained in the baseline models.

5.3 ESG pillar composition

In Figure 3, ESG pillar composition, Governance disclosure represents 75.5% of all ESG-classified sentences, which can be interpreted as the major contributing category to ESG disclosure, while social and environmental disclosure account for 12.8 and 11.7%, respectively. It should also not be considered as the selected sample of companies have better governance performance, rather governance topics are higher in the prospectus compared to others. Also, it is to be noted that companies are giving less attention to disclosures related to environmental and social factors; this imbalance should be noted by the regulators.

5.4 Temporal evolution of ESG disclosure

Figure 4 shows the temporal evolution of ESG as well as E, S, and G disclosure scores. The mean overall ESG disclosure score shows an improving trend over the study period, from roughly 11.1% in FY2021-22 to about 11.7% in FY2025-26, with only a small decline in FY2022-23 followed by continued growth. Governance remains the dominant pillar every year and shows minimal increase throughout the period. The environmental score also shows a similar trend of improving over the year, whereas the social score is comparatively stable. The temporal evolution also resembles the pillar composition result, and it can be expected to continue as during the period, the gap never reduced between governance and the other two categories.

TABLE 3 Descriptive statistics for prospectus-level ESG disclosure measures.

Variable	*N*	Mean	Median	Std. dev.	Min	Q1	Q3	Max
Total clean sentences	310	6365.323	6146.000	1408.080	3870.000	5411.500	7112.000	13160.000
ESG sentence count	310	719.713	705.500	159.747	361.000	606.000	814.500	1398.000
Environmental disclosure score (%)	310	1.314	1.132	0.961	0.077	0.681	1.697	7.645
Social disclosure score (%)	310	1.462	1.398	0.516	0.493	1.132	1.686	4.846
Governance disclosure score (%)	310	8.605	8.692	1.001	5.783	7.869	9.235	11.596
Overall ESG disclosure score (%)	310	11.382	11.334	1.393	7.860	10.383	12.202	17.772

TABLE 4 Descriptive statistics for supplementary firm characteristics.

Variable	*N*	Mean	Median	Sd	Min	Max
Board Size	310	7.558065	8	1.697566	5	14
Board Independence	310	0.494331	0.5	0.078286	0	0.833333
ln (Issue Size)	310	6.583381	6.476972	1.076761	4.094345	10.23491
Firm Age	310	20.77742	18	12.14258	1	68
ROA	310	0.086469	0.078106	0.086982	−0.17421	0.303246
ROE	310	0.17723	0.187291	0.198238	−0.61953	0.710881
D/E Ratio	310	0.645192	0.443724	0.780301	0	4.228462
ln (Total Assets)	310	6.739854	6.628971	1.285331	3.644286	12.12812
ln (Sentence Count)	310	8.735697	8.723557	0.212154	8.26101	9.484937

5.2 Correlations and Multicollinearity Diagnostics

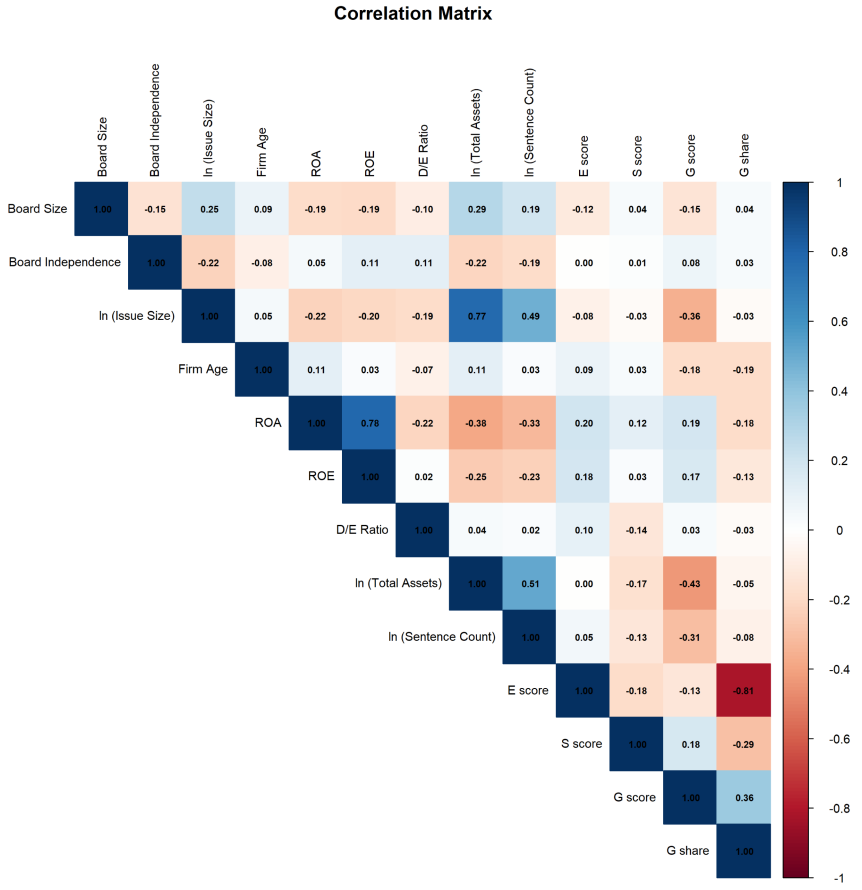


FIGURE 2
Correlation matrix of supplementary firm characteristics.

5.5 Sectoral heterogeneity

Figure 5 shows the average overall ESG disclosure score of the different sectors in the study. Average ESG disclosure differs across sectors. Utilities and commodities are leading the list, while telecommunications records the lowest sector average of 9.44%, but it is also due to the smaller number of companies in the sector, so further data and results are required for generalizing it. Most of the sectors are around approximately 11–12% of ESG disclosure, indicating less deviation from the overall ESG disclosure, even though there are sector-specific business models and regulatory exposures that produce meaningful differences.

Figure 6 present sector wise E, S and G disclosure score. When we dig deeper into the pillar-level disclosure of sectors, the results are still similar, where Governance disclosure leads in every sector, confirming that the aggregate result is not driven by a single

TABLE 5 VIF results for supplementary firm characteristics.

Variable	VIF
`Board Size`	1.135
`Board Independence`	1.099
`ln (Issue Size)`	2.973
`Firm Age`	1.077
ROA	3.596
ROE	3.001
`D/E Ratio`	1.296
`ln (Total Assets)`	3.237
`ln (Sentence Count)`	1.467

industry; all industries together contribute to it. Environmental disclosure is relatively higher in the utilities sector, 4.56, which is far higher when we compare with other sectors, while social disclosure shows a lower disclosure across sectors. The heat map shows that sectoral heterogeneity operates primarily around a common governance-dominant baseline.

The mean ESG disclosure intensity for utilities is the highest at 13.497%, followed by commodities at 12.194% (Figure 7). The distributions of the other sectors are fairly similar and are clustered around 11%, while telecommunications have the lowest mean score of 9.441%. The distribution is also the widest for utilities and the tightest for energy. There is a high level of overlap in most sectors, suggesting that disclosure intensity differs significantly within sectors and is not solely based on sector affiliation. The small sample size for telecommunications, energy, and utilities results in a cautionary interpretation of the results for these industries.

5.6 ESG topic and subcategory-level disclosure patterns

However, there is also more detailed insight provided by the subcategory heat map (Figure 8) concerning the governance-dominated theme. Governance-themed subcategories make the darkest and most persistent cluster within the sectors, even though there is a difference in their exact composition. Environmental categories cluster in those sectors that have the most exposure to natural resources, pollution, water and ecosystems, whereas social categories reveal sector peaks selectively.

Topic emphases change over time in a different way for various industries (Figure 9A). Topics related to governance persist across all sector panels; however, their proportions vary over time. Water emerges as an important topic in some environmentally sensitive sectors, whereas disclosure, transparency, and governance topics fluctuate over time. This confirms the application of topic shares by sector-year instead of the pooled ranking of topics.

The chosen sub-categories are the top 10 categories under the main E, S, and G pillars that are exhibiting different trends (Figure 9B). Climate change disclosure increases in later years, while human rights and core labor standards decrease in later years. Corporate governance disclosure is fairly constant, while sustainability management and reporting increase early on but later decrease. These reflect the changing prominence of disclosures, but not of changes in ESG performance.

5.7 Robustness and diagnostic analysis

The scatter plot (Figure 10) indicates no mechanically rising trend of the ESG disclosure score with the increasing length of prospectuses. There is a slight negative relationship found between the ESG score and the length of prospectuses, suggesting that longer documents may contain proportionally more non-ESG material even when they contain many ESG sentences in absolute terms. This diagnostic supports the use of a normalized score.

5.8 Regression analysis

The conditional estimates based on the OLS regression are presented in Figure 11, which shows how different sectors and different times affect the ESG disclosure score compared to the Consumer Discretionary sector and the baseline year of 2021–22. It can be seen from the sector estimates that there are significant differences in terms of how these sectors report their performance; Utilities, Commodities, and Healthcare have significantly higher levels of disclosure compared to the reference group, while the Telecommunication sector has a significant negative effect. The confidence interval for the coefficients for other sectors (Energy, Industrials, and Fast Moving Consumer Goods) intersects with zero, which means that there is no statistical difference in their reporting practice. From the perspective of the longitudinal effects, the year fixed effects show the increasing trend with the highest level and significance achieved in the fiscal year 2025–26. It should be noted that all these associations do not establish causality (Figure 11).

The OLS estimates (Table 6) show significant sectoral and temporal differences in overall ESG disclosure intensity. Relative to Consumer Discretionary, Utilities ($\beta = 2.437$, $p < 0.001$), Commodities ($\beta = 1.186$, $p < 0.001$), Healthcare ($\beta = 0.594$, $p = 0.019$), and Industrials ($\beta = 0.464$, $p = 0.024$) have significantly higher disclosure scores, whereas Telecommunications have a significantly lower score ($\beta = -1.716$, $p = 0.026$). Energy, FMCG, Information Technology, and Services do not differ significantly from the reference sector. Relative to FY2021–22, only FY2025–26 shows a statistically significant increase in ESG disclosure intensity ($\beta = 0.741$, $p = 0.002$). The results indicate conditional sector and year differences and do not establish causal effects or differences in underlying ESG performance.

5.9 Firm characteristics, board structure and ESG disclosure composition

Table 7 examines whether observable issuer characteristics and board structure are associated with the intensity and composition of ESG disclosure. The four specifications are jointly significant, although their explanatory power varies across the ESG dimensions, with R^2 values ranging from 0.225 for social disclosure to 0.513 for environmental disclosure. This variation itself suggests that the determinants of prospectus ESG communication are not uniform across environmental, social, and governance themes.

For governance disclosure, total assets are negatively associated with the G score ($\beta = -0.266$, $p < 0.01$), while issue size is not statistically significant. Prospectus length is also negatively associated with the G score ($\beta = -1.060$, $p < 0.01$), indicating that longer prospectuses do not necessarily devote a greater proportion of their content to governance-related disclosure. Firm age shows only a weak negative association. In the environmental model, board

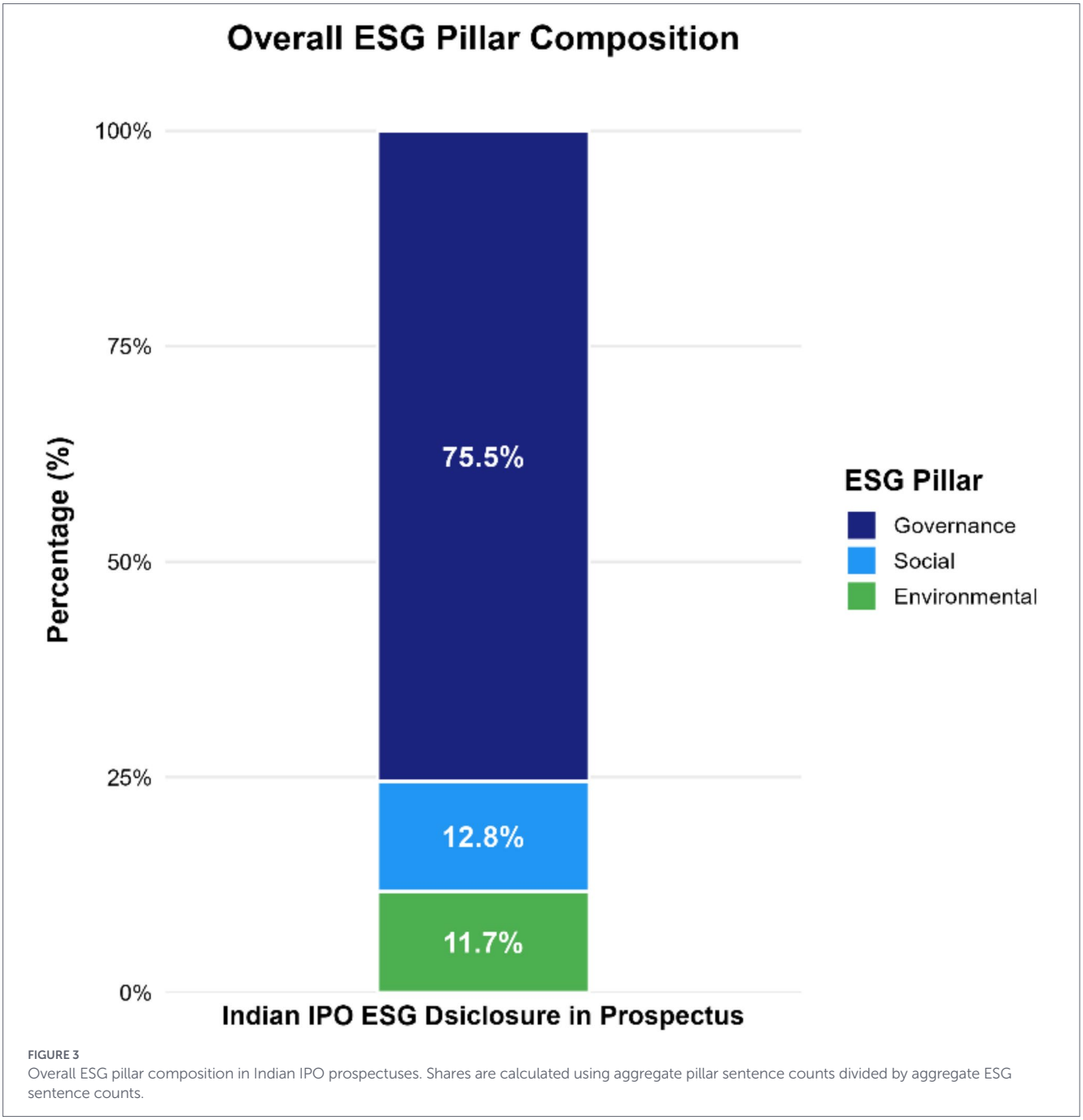


FIGURE 3
Overall ESG pillar composition in Indian IPO prospectuses. Shares are calculated using aggregate pillar sentence counts divided by aggregate ESG sentence counts.

independence is negatively associated with the E score ($\beta = -1.169$, $p < 0.05$), while ROA and leverage show only marginal associations. The social model provides comparatively limited evidence of systematic relationships with the examined firm and board characteristics, with no predictor significant at the 5% level. For governance share, firm age is negatively associated with the relative emphasis placed on governance content ($\beta = -0.001$, $p < 0.01$), whereas the positive association with board independence is only marginally significant.

Taken together, the results do not support a simple pattern in which larger, more profitable, or otherwise observably stronger issuers uniformly disclose more ESG information. Instead, the associations differ across the three ESG pillars and between disclosure intensity and disclosure composition. This pattern is consistent with the possibility that differentiation at the IPO stage occurs partly through the allocation and emphasis of ESG communication rather than through disclosure volume alone. However, the estimates represent conditional associations and should not be interpreted as evidence that the examined firm characteristics causally determine disclosure strategy or underlying ESG quality.

Table 8 examines whether board structure is associated with governance-oriented disclosure in IPO prospectuses. Board Size has a negative coefficient ($\beta = -0.072$, $p = 0.053$), while Board Independence has a positive but statistically insignificant coefficient ($\beta = 0.953$, $p = 0.152$). Although neither variable is individually significant at the 5% level, the HC3-robust joint Wald test rejects the null hypothesis that the coefficients on Board Size and Board Independence are simultaneously zero [$F(2, 294) = 3.880$, $p = 0.022$]. This indicates that the two board characteristics, when considered

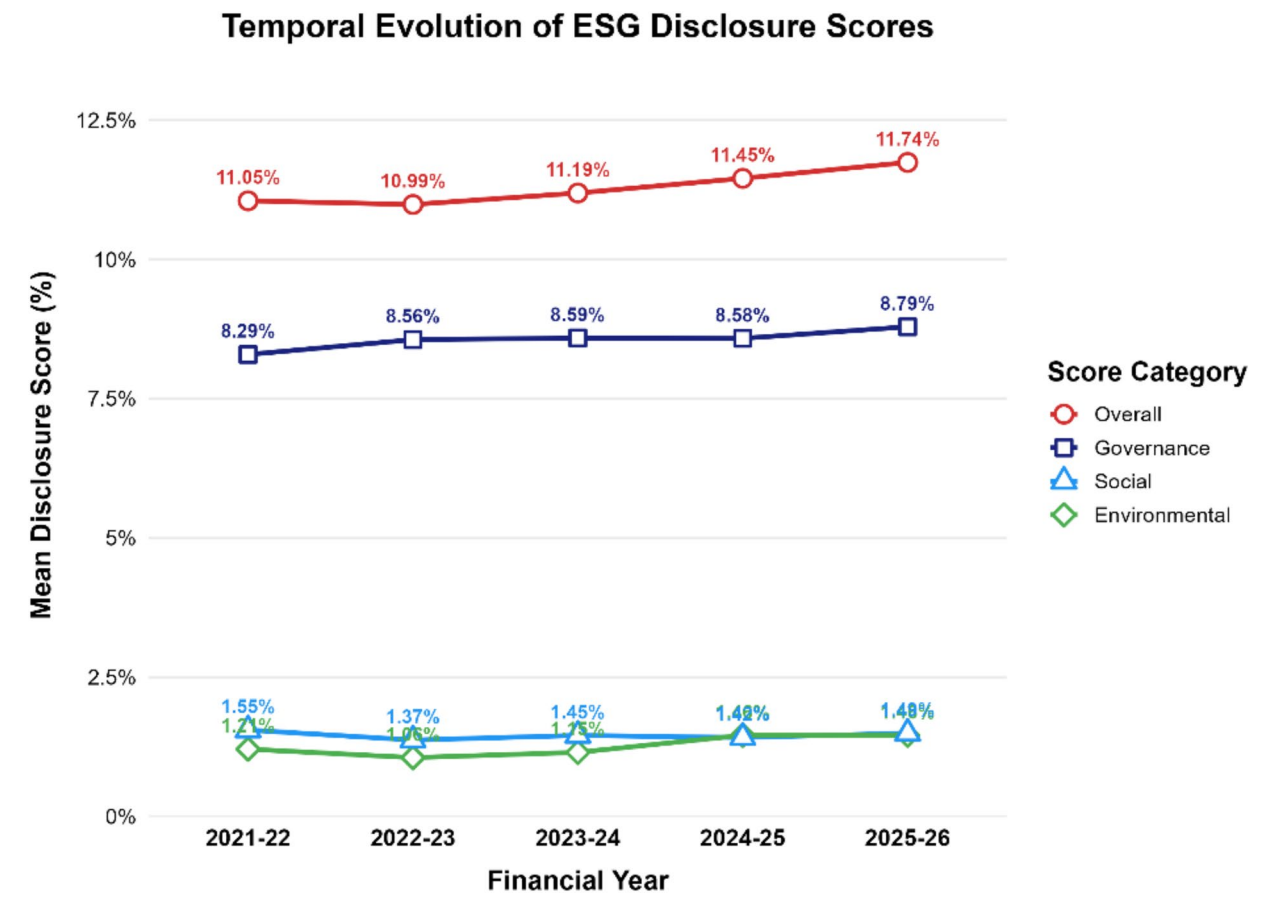


FIGURE 4
Temporal evolution of mean ESG disclosure scores, FY2021-22 to FY2025-26. Scores are percentages of total clean prospectus sentences and are averaged within each financial year.

jointly, are associated with variation in governance disclosure after controlling for sector and year effects. The result provides qualified evidence consistent with a credibility-related signaling channel. Governance structures are relatively observable characteristics of an issuer and may therefore contribute to how governance readiness is communicated at the IPO stage. However, the individual coefficients are not statistically significant at the 5% level, and the model explains only a modest share of the variation in governance disclosure ($R^2 = 0.091$). The findings should therefore be interpreted as a joint association between board structure and governance communication rather than as evidence that either larger boards or greater board independence independently improves disclosure credibility. As the analysis is observational, the results do not establish a causal effect of board structure on governance disclosure.

6 Discussion

The current research sought to assess the level and composition of ESG disclosures in the prospectuses of 310 Indian main board IPOs made in the fiscal years 2021–22 through 2025–26, employing semantic classification of sentences. Three major insights can be derived. ESG sentences constituted a relatively small portion of the total number of sentences in the prospectuses (11.38%), with governance being the largest portion of such sentences (75.5%) as opposed to the social (12.8%) and environmental (11.7%) pillars of ESG. The level of overall ESG disclosure did not grow significantly, with a statistically distinguishable increase seen only in the financial year 2025–26 and not as a consistent annual trend. The disclosure also differed across sectors, although the governance continued to prevail in every sector. Therefore, this paper offers not just another ESG score but also its composition.

6.1 Governance dominance and the meaning of aggregate ESG disclosure

The governance dimension suggests that the Indian use of ESG-related wording in offer documents centers around board composition, audit, internal control, compliance, ethics, and shareholders' protection. This observation corroborates the results of Baier et al. (2020), showing that terms related to corporate governance dominate the ESG lexicon in American 10-K filings and proxy statements, at least partially due to the fact that governance issues are mandatory disclosures. At the same time, this observation is consistent with Bae et al. (2018), who show that the association between corporate sustainability disclosure and board characteristics and ownership in South Asia can be explained via the signaling channel. The fact is that in an IPO, the information asymmetry between issuer and investors is particularly high, while the governance-related information is relatively easy to verify and is already contained in law-based documents.

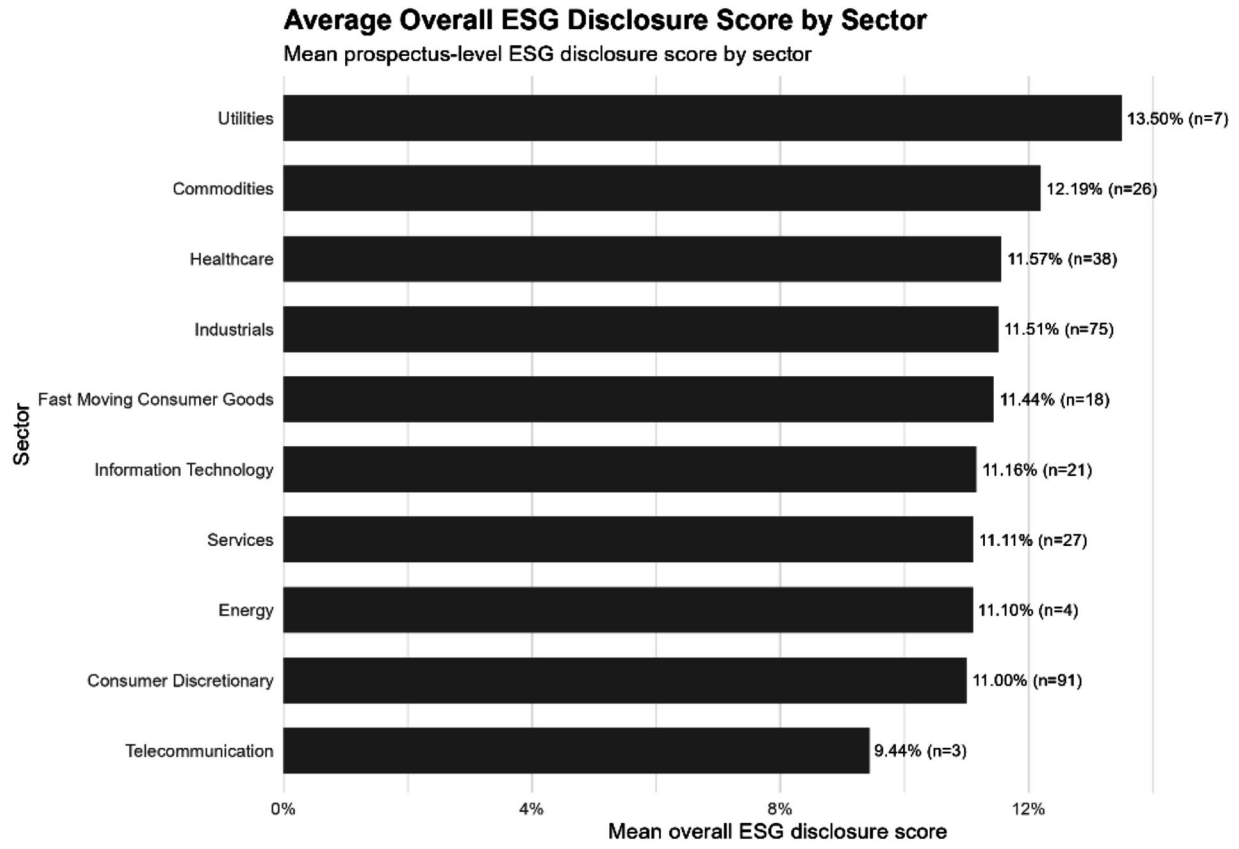


FIGURE 5
Average overall ESG disclosure score by sector. Bars show mean prospectus-level overall ESG disclosure scores; sector sample sizes are displayed in the figure.

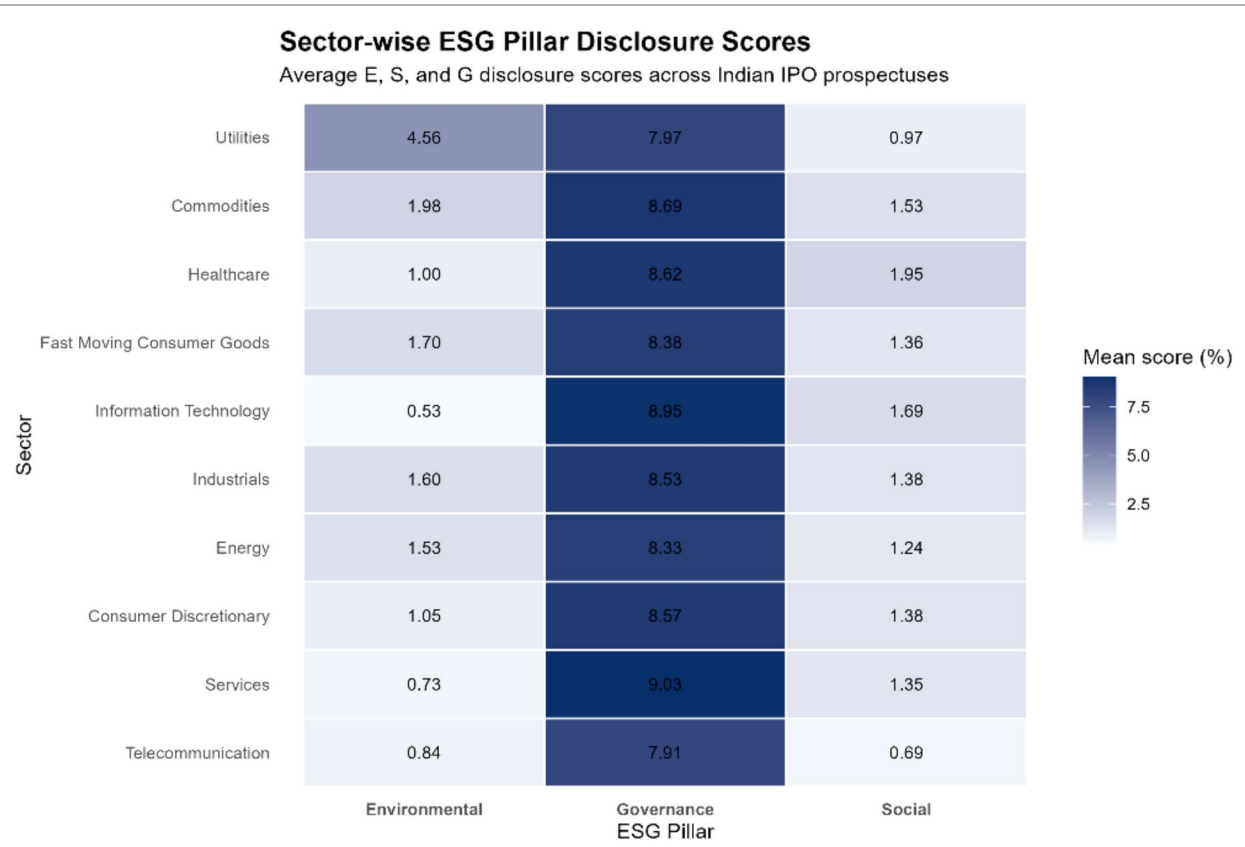


FIGURE 6
Sector-wise mean ESG pillar disclosure scores. Each cell reports the mean pillar score as a percentage of total clean prospectus sentences within the sector.

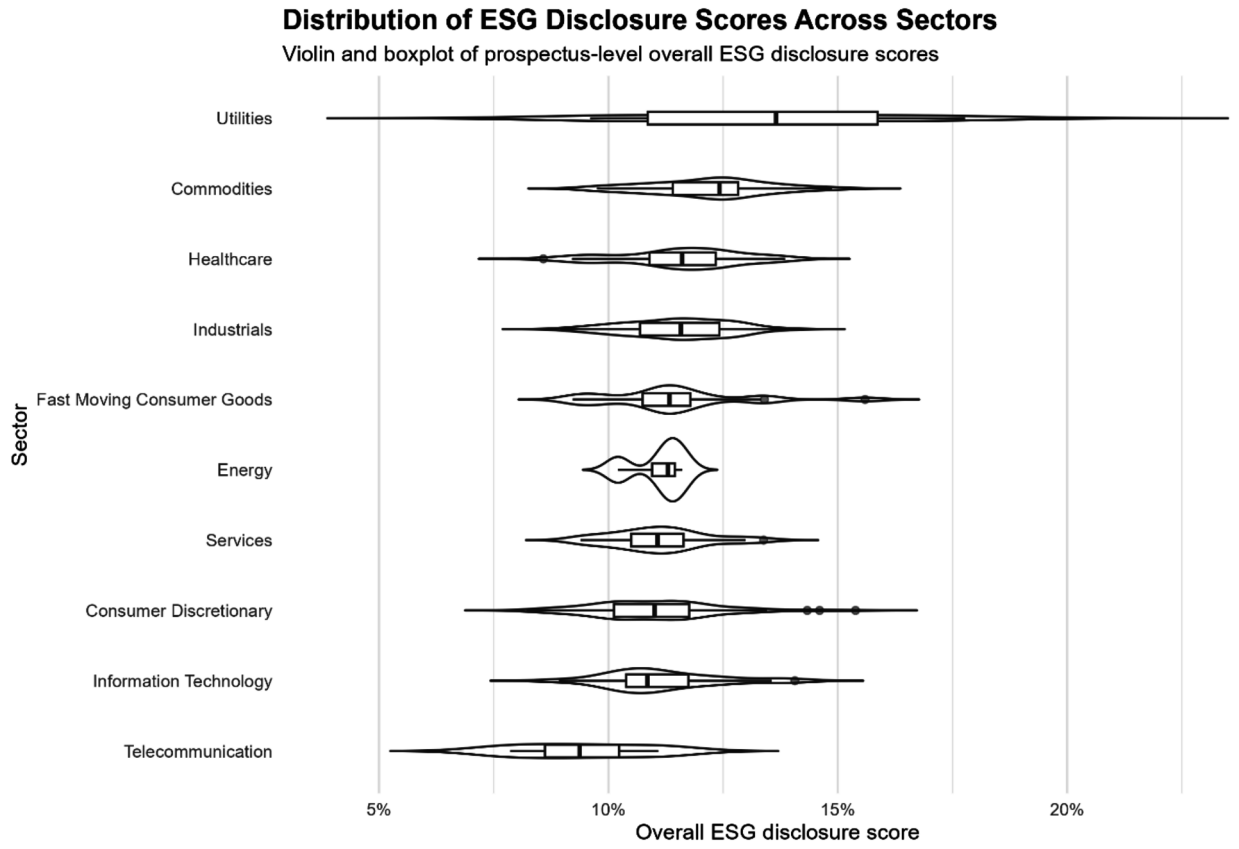


FIGURE 7
Distribution of prospectus-level overall ESG disclosure scores across sectors. Violin densities and embedded box plots show within-sector variation in overall ESG disclosure intensity.

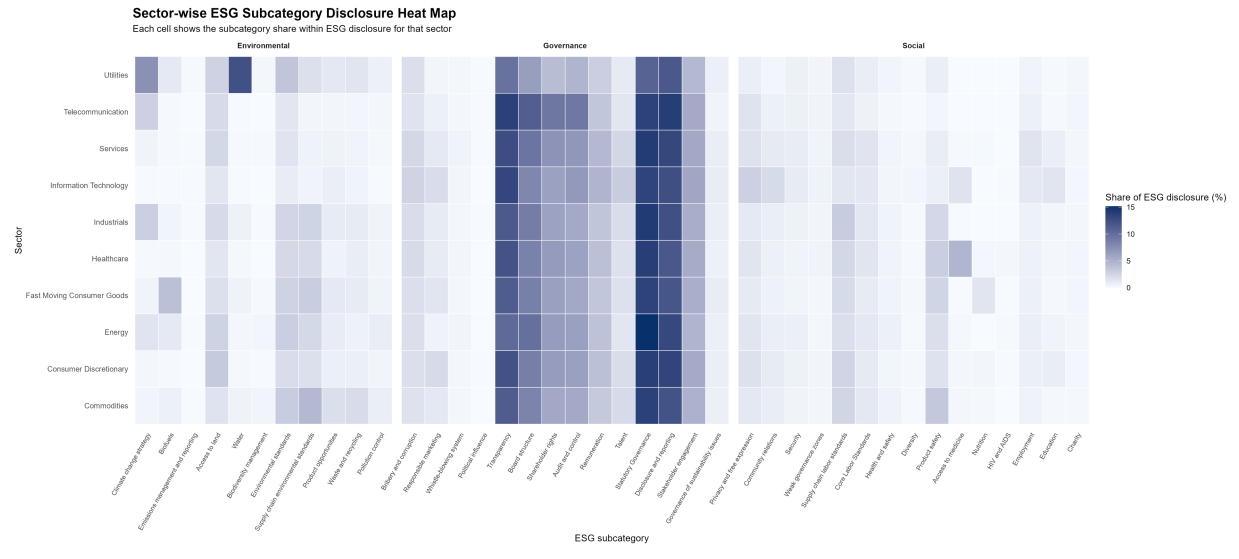


FIGURE 8
Sector-wise ESG subcategory disclosure heat map.

This result qualifies, rather than contradicts, prior IPO evidence. According to Kirubakaran and Joseph (2026) aggregate ESG disclosure in Indian IPO prospectuses is positively linked to listing performance, whereas according to Abbas et al. (2025) community and environmental disclosure extent and quality impact the long-term performance of Malaysian IPOs. Based on the above decomposition, an aggregate ESG score of Indian IPOs could be considered mainly as a governance signal unless separate pillar weights are given. Thus, having a high aggregate ESG score does not mean that the firm possesses both strong social and environmental characteristics. The reason for the differentiation is that empirical evidence shows that the level of disclosure and its quality

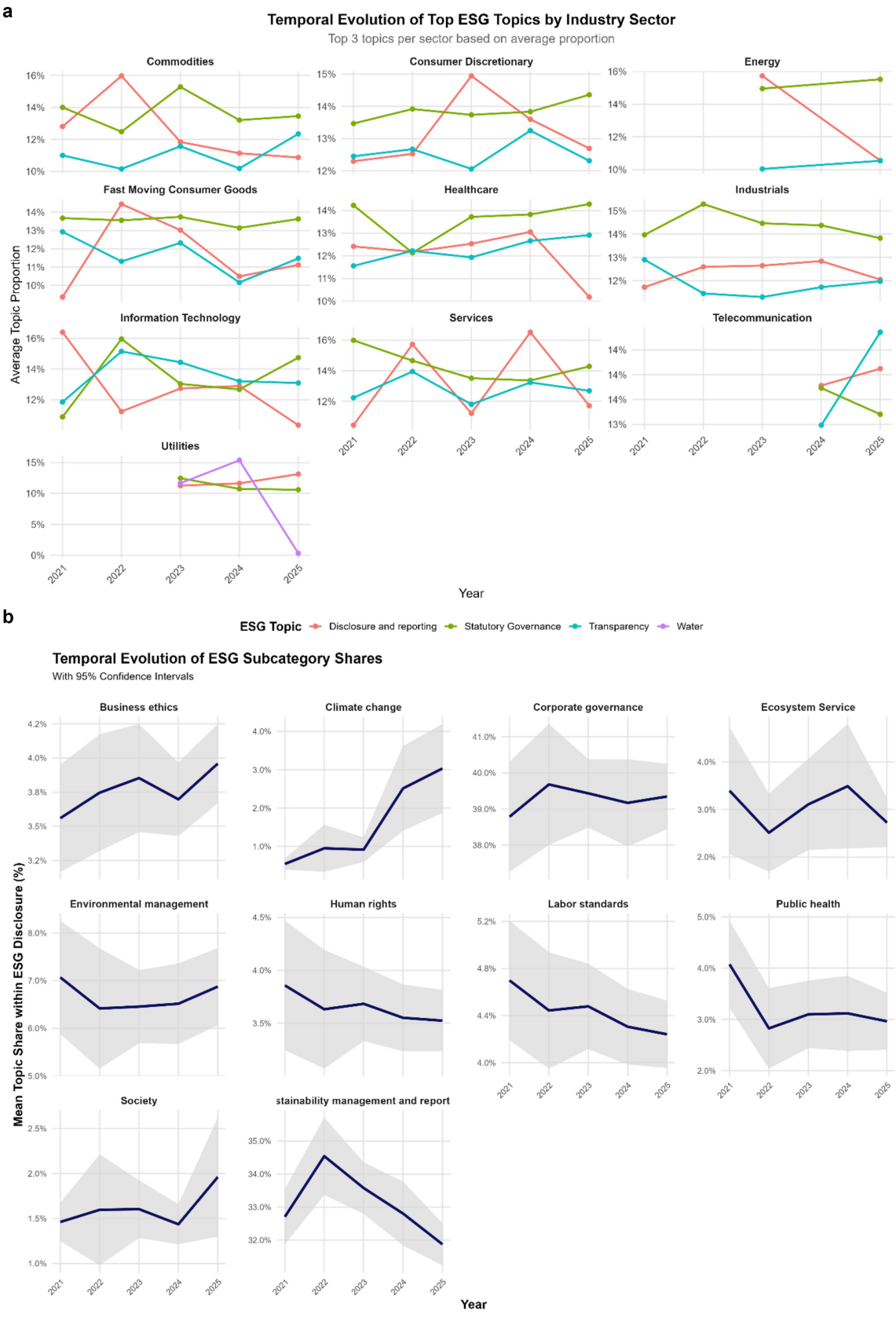


FIGURE 9
(a) Temporal evolution of top ESG topics by industry sector. **(b)** Temporal evolution of selected ESG subcategory shares. Shaded bands show 95% confidence intervals. The vertical axis is the mean topic share within ESG disclosure.

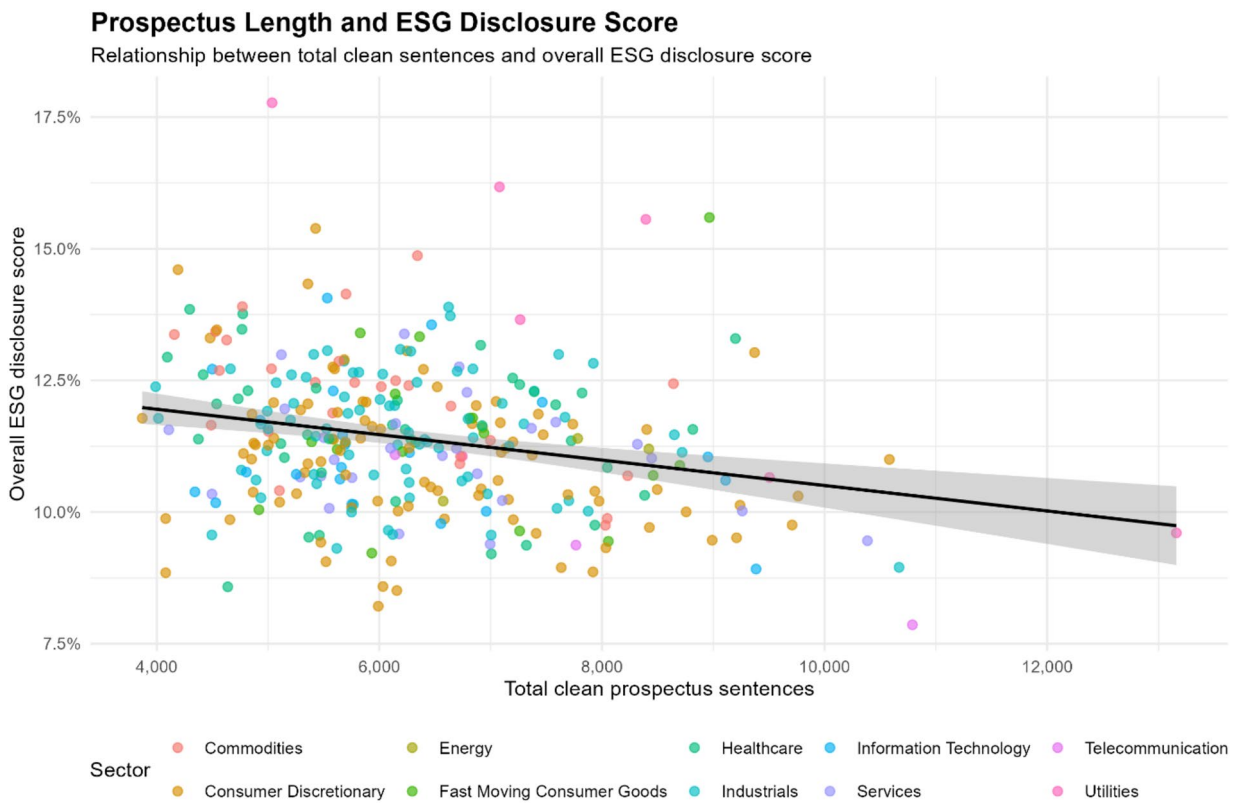


FIGURE 10
Prospectus length and overall ESG disclosure score.

cannot be used interchangeably. As noted by Hummel and Schlick (2016), high sustainability performers disclose better compared to weak performers that use poor quality disclosure to achieve legitimacy. Michelon et al. (2015) demonstrate the same results, indicating that using common practice does not necessarily lead to more accurate, balanced, and substance-based information. On the contrary, Arena et al. (2015) find that under certain governance conditions, positive environmental language could be considered informative with respect to the future environmental performance of the firm.

The supplementary regressions complicate a simple differentiation narrative. Larger issuers (by total assets) disclose significantly less governance-related content (Table 7, $\beta = -0.266$, $p < 0.01$), which runs counter to a straightforward reading in which observably larger or more established firms use governance disclosure to signal superior quality. One plausible explanation is compensatory rather than purely differentiating signalling: smaller or less-established issuers, lacking alternative track-record-based signals such as public trading history, analyst coverage, or brand recognition, may lean more heavily on the one channel most readily available and most credible at the pre-listing stage, verifiable governance structure. Under this reading, governance-disclosure intensity reflects the relative importance of the signal to a given issuer rather than a marker of quality per se. This interpretation is consistent with the joint, but individually modest, association between board structure and governance disclosure reported in Table 8, and underscores that the present analysis characterizes patterns of association rather than establishing which mechanism differentiation, compensation, or regulatory coercion dominates.

6.2 Temporal evolution and regulatory diffusion

While the increase in β from around 11.1% in FY2021-22 to 11.7% in FY2025-26 is relatively small in economic terms, only the last-year coefficient is statistically significantly different from the baseline ($\beta = 0.741$, $p = 0.002$). The late-increase trend thus appears as a late-window phenomenon rather than annual growth. A possible reason might be delayed institutional adaptation: corporate boards, legal counsel, and merchant bankers need some time to move their post-listing ESG reporting experience into pre-listing drafting practices. Consistent evidence from Korea suggests that ESG rhetoric developed in stages, rather than consistently, with recent ESG reports paying increasingly greater attention to carbon neutrality, supply chain management, risk, and compliance issues (Lee et al., 2026). In addition, cross-country quasi-experimental evidence points to positive impacts of mandatory ESG disclosure on the price informativeness for firm-level information (Zhang et al., 2023), and the EU disclosure mandate had differential impacts on the markets, dependent on pre-disclosure ESG reporting behavior and performance of the firm (Grewal et al., 2019). This all makes regulatory diffusion an acceptable explanation for the late increase, yet the current pooled analysis cannot single out an Indian intervention responsible for that increase.

The motivations of investors might accentuate this institutional mechanism. According to Mulchandani et al. (2022) ESG disclosures reduce the cost of equity capital for Indian listed companies, which provides a motivation for potential issuers to

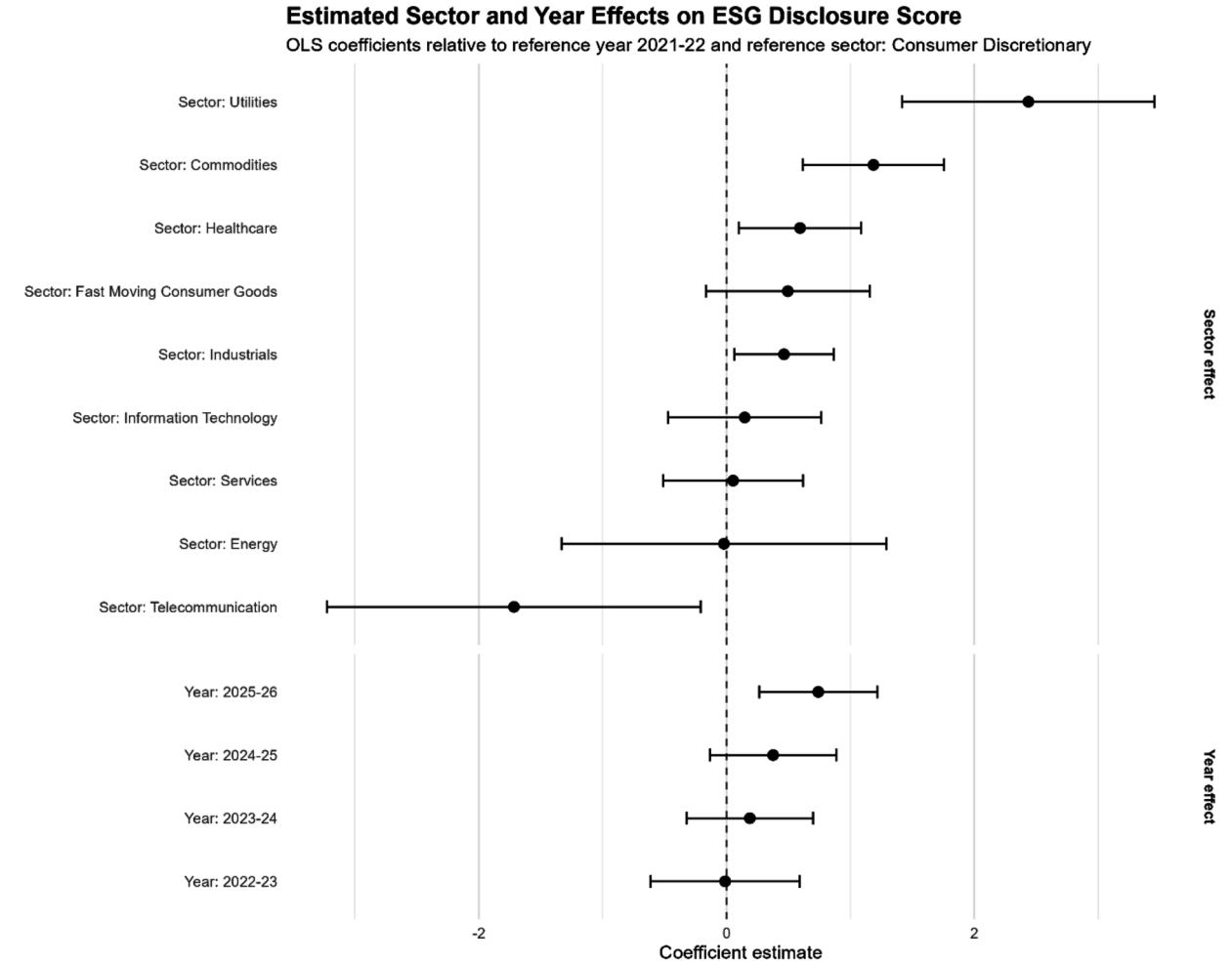


FIGURE 11
Estimated sector and year effects on the overall ESG disclosure score. OLS coefficients are shown relative to Consumer Discretionary and FY2021-22; horizontal lines are 95% confidence intervals.

disclose more ESG information. However, it would be a mistake to interpret the insignificance of the coefficients from the intervening years as evidence of the lack of any change in disclosure behavior. Heterogeneous adoption, low precision of measures in annual subsamples, and variations in substance not captured by overall sentence shares are all possible explanations for these results.

6.3 Sectoral heterogeneity around a common disclosure baseline

Adjusting for the influence of year dummies, Utilities, Commodities, Healthcare, and Industrials have significantly higher ESG disclosure intensity than Consumer Discretionary; Telecommunications have significantly lower intensity, while Energy, Fast Moving Consumer Goods, Information Technology, and Services do not differ from the reference sector significantly. Utilities also have the highest share of environmental disclosures (4.56%). This is in line with a materiality perspective: sectors that experience visibility of emissions, resource consumption, land, water, safety, or health-related matters have strong incentives to disclose these issues within offer documents. In accordance with this perspective, Lee et al. (2026) note that the ESG discourse of Korea demonstrates significant sector-specific differences, where energy and finance disclose matters of climate and governance, construction focuses on safety issues, while manufacturing discloses matters of environmental management. Abbas et al. (2025) demonstrate that the market significance of environmental and community disclosure differs depending on the disclosure dimension.

Sector is not, however, able to disrupt the prevailing order of disclosure, as governance still constitutes the biggest pillar in all sectors. Hence, the implications of these results point to differentiated disclosure intensity in a similar compliance-centric framework, rather than separate ESG disclosure frameworks in different sectors. The estimate for Telecommunications must be viewed very carefully, since the small sector sample may cause issuer effects to become very important; this is also true for the most dispersed distributions in Energy and Utilities.

6.4 Methodological interpretation

The lack of any positive correlation between the length of the prospectus and the normalized ESG score suggests the use of

TABLE 6 Regression coefficients.

Variable	Coefficient	Std. error	t-statistic	*p*-value	95% Confidence interval
Constant/Intercept	10.659***	0.226	47.249	<0.001	[10.215, 11.103]
Year: 2022–23	−0.011	0.306	−0.037	0.971	[−0.613, 0.591]
Year: 2023–24	0.188	0.259	0.728	0.467	[−0.321, 0.698]
Year: 2024–25	0.376	0.259	1.452	0.147	[−0.134, 0.886]
Year: 2025–26	0.741***	0.242	3.06	0.002	[0.264, 1.217]
Sector: Commodities	1.186***	0.289	4.099	<0.001	[0.617, 1.756]
Sector: Energy	−0.021	0.666	−0.031	0.975	[−1.332, 1.290]
Sector: Fast Moving Consumer Goods	0.495	0.336	1.475	0.141	[−0.165, 1.155]
Sector: Healthcare	0.594**	0.251	2.362	0.019	[0.099, 1.088]
Sector: Industrials	0.464**	0.204	2.275	0.024	[0.063, 0.865]
Sector: Information Technology	0.147	0.314	0.468	0.64	[−0.472, 0.766]
Sector: Services	0.053	0.287	0.185	0.853	[−0.512, 0.618]
Sector: Telecommunication	−1.716**	0.766	−2.24	0.026	[−3.223, −0.208]
Sector: Utilities	2.437***	0.517	4.711	<0.001	[1.419, 3.455]

Statistic	Value
Number of observations	310
R-squared	0.174
Adjusted R-squared	0.138
F-statistic	4.799
Model p-value	<0.001
Residual standard error	1.296
Degrees of freedom	296

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

sentences about ESG as a proportion of clean sentences, instead of simply using their number. Longer prospectuses seem to add more legal, financial, and risk-factor text without necessarily adding any ESG intensity. The semantic method also builds upon previous dictionary methods by adding the meaning categorization at the sentence level. Baier et al. (2020) prove the scalability of the transparent ESG taxonomy system, and Menon et al. (2026) prove the use of Sentence-BERT to measure the disclosure at the pillar level in the annual reports of India. However, such semantic classification techniques measure the discourse on ESG, rather than the actual quality or performance of ESG policies. According to Lee et al. (2026) textual data is only interpretively justified as discourse until it is related to some operations or external validation.

7 Implications, limitations, and future research

The structure of this research highlights its practical regulatory and investment relevance. Regulators and stock exchanges may consider how the guidance for offer documents might benefit from requiring a clear distinction between environmental, social, and governance data to be presented, including relevant industry metrics, material risks, targets, time horizon, and the rationale behind forward-looking claims. There is no evidence supporting the imposition of a standard volume measure; rather, sector-based findings imply that materiality-based expectations should be applied. Institutional investors, underwriters, and analysts need to be careful about using a one-size-fits-all ESG percentage in relation to sustainability performance. Pillar-specific percentages, specificity of reporting, quantifiable commitment, and external assurance should be considered independently because of their financial significance, even when they are compositionally governed (Kirubakaran and Joseph, 2026; Mulchandani et al., 2022).

The study also has implications for methodology. Sentence embeddings permit the efficient evaluation of large prospectus corpora, where manual coding is difficult, and might facilitate pre-filing diagnostics for pillar balance. Yet, any practical application requires further validation beyond the current methodology. The classifier uses only one model for the embeddings along with a 0.45 threshold of similarity, which was determined from a small number of manually-coded sentences rather than an independent gold-standard coding scheme. For future studies, different embedding models and similarity thresholds must be tested, and results should be presented in terms of precision, recall, F1 score, and confusion matrix, in addition to inter-coder reliability against a stratified human coding. In addition, a taxonomy built solely based on India's Business Responsibility and Sustainability Reporting framework and relevant vocabulary should be created.

TABLE 7 Firm quality, board structure, and ESG disclosure composition.

Variable	(1) G score	(2) E score	(3) S score	(4) G share
ln (issue size)	0.009	−0.071	0.065	0.001
	(0.089)	(0.069)	(0.046)	(0.005)
Firm age	−0.009*	0.007	0.001	−0.001***
	(0.005)	(0.004)	(0.002)	(0.000)
ln (total assets)	−0.266***	0.079	−0.077*	−0.007
	(0.080)	(0.058)	(0.040)	(0.004)
ROA	−0.025	1.596*	−0.256	−0.096
	(1.020)	(0.928)	(0.610)	(0.067)
ROE	0.333	0.018	0.016	0.008
	(0.425)	(0.359)	(0.243)	(0.027)
D/E ratio	0.060	0.126*	−0.053	−0.004
	(0.072)	(0.066)	(0.040)	(0.004)
Board size	0.027	−0.028	0.019	0.001
	(0.034)	(0.023)	(0.018)	(0.002)
Board independence	−0.216	−1.169**	−0.051	0.072*
	(0.612)	(0.526)	(0.376)	(0.041)
ln (sentence count)	−1.060***	−0.188	−0.240	−0.002
	(0.363)	(0.269)	(0.176)	(0.019)
Constant	19.126***	3.841*	3.577**	0.743***
	(2.958)	(2.194)	(1.464)	(0.160)
Sector fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations (N)	310	310	310	310
R-squared	0.299	0.513	0.225	0.413
Adjusted R-squared	0.245	0.476	0.166	0.368
F-statistic	6.507	12.368	3.787	9.169
F-statistic *p*-value	<0.001	<0.001	<0.001	<0.001
Breusch-Pagan *p*-value	0.004	<0.001	0.951	0.314
Maximum VIF	3.902	3.902	3.902	3.902
Standard errors	HC3 Robust	HC3 Robust	Conventional OLS	Conventional OLS

OLS estimates are reported with standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$. Sector and year fixed effects are included in all specifications. The Breusch-Pagan test is conducted separately for each model. HC3 heteroskedasticity-robust standard errors are reported where the Breusch-Pagan test rejects homoskedasticity at the 5% level. ROA, ROE, and D/E Ratio are winsorized at the 1st and 99th percentiles. Issue Size, Total Assets, and Sentence Count are natural-log transformed. Maximum VIF refers to the full specification used in the corresponding model.

A number of restrictions affect interpretability. First, the pooled OLS model using controls for sector and year provides only a conditional association, not causality. For example, an improved approach could leverage the phased rollout of Indian ESG regulations by using differences-in-differences, interrupted time series, and event study approaches where the treatment and control groups are unambiguously delineated. The goal here is to assess whether there were any discrete jumps in the ESG reporting content in reaction to regulatory milestones or whether the increase in the level of ESG reporting was just part of a broader trend. Second, the reporting index measures the intensity of text without addressing the questions of specificity, balance, assurances, and implementation. Future research needs to develop quality indicators for quantified targets, baseline years, capital investment, outcome measurements, negative disclosures, external assurances, and consistency between prospectuses, BRSR filings, and actual ESG performance. This is important since the existing literature demonstrates that the possibility of a high volume of disclosure can coexist with symbolic or low-quality reporting (Hummel and Schlick, 2016; Michelon et al., 2015).

Comparative analyses can explore whether governance dominance is peculiar to the Indian prospectus structure or a feature of emerging economies in transition to mandatory ESG reporting. The reason behind the dominance of governance related content might be due to SEBI ICDR provisions on the contrary environmental and social contents are not pushed by any provision for IPOs in India but BRSR on the other side is mandatory for top companies, it would be more investor friendly especially SRI friendly if the regulators force the companies to disclose more about environmental and social activities of the companies.

Lastly, future research will investigate the relationship between the structure of the pillar and the quality of disclosure with IPO underpricing, subscription behavior, analyst coverage, cost of capital after listing, litigation risk, and eventual environmental or social outcomes. This will help determine whether governance dominance is just the most frequent type of ESG reporting or also the most valued one.

TABLE 8 Board structure and governance disclosure: credibility-specific model.

Variable	Model 5: G score
Board size	−0.072*
	(0.037)
Board independence	0.953
	(0.663)
Constant	8.437***
	(0.495)
Sector fixed effects	Yes
Year fixed effects	Yes
Observations (*N*)	310
R-squared	0.091
Adjusted R-squared	0.044
Overall F-statistic	2.042
Overall F-statistic *p*-value	0.013
Breusch-Pagan *p*-value	0.004
Maximum VIF/GVIF	1.302
Standard errors	HC3 Robust
Joint Board Wald F-test	3.880 [$F(2, 294)$]
Joint Board Wald *p*-value	0.022

The dependent variable is the G score. OLS coefficient estimates are reported with standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Sector and year fixed effects are included. Because the Breusch-Pagan test indicates heteroskedasticity, HC3 heteroskedasticity-robust inference is reported. The joint Board Wald test evaluates the null hypothesis that the coefficients on Board Size and Board Independence are simultaneously zero.

8 Conclusion

The study offers a sentence-level analysis of the disclosure of ESG in 310 Indian main board IPO prospectuses released from financial years 2021–22 to financial years 2025–26. The ESG content represents a relatively small share of the total prospectus text and is structurally clustered in the governance domain, while the disclosure of the environmental and social areas remains relatively weak. The overall intensity shows a marginal increase, with a marked difference emerging in the last year, with sector-specific differences, but not enough to take away the leading role of the governance pillar. The normalized score is not automatically driven by the length of the prospectus, thereby justifying its use as a comparative metric of text intensity.

Beyond this descriptive pattern, the supplementary firm-level analysis (RQ4) indicates that observable issuer characteristics are not uniformly related to ESG disclosure across pillars. Notably, larger issuers by total assets disclose significantly *less* governance-related content, a result that qualifies rather than confirms a simple narrative in which larger or more established firms straightforwardly differentiate themselves through greater governance disclosure. Board size and board independence are jointly, though not individually, associated with governance disclosure intensity, providing qualified rather than definitive support for a credibility-related signaling channel grounded in verifiable board structure.

The primary contribution of this research is twofold. First, the composition of ESG disclosure in the IPO process differs meaningfully from its overall volume: for Indian IPO prospectuses, an aggregate ESG signal is predominantly a governance signal, and the balance and quality of disclosure should be assessed before drawing inferences about environmental and social commitments. Second, the supplementary analysis provides an initial, necessarily qualified empirical bridge between this compositional pattern and the signalling and credibility mechanisms that motivate it, showing that issuer and board characteristics are associated with disclosure in pillar-specific rather than uniform ways. These observations are descriptive and correlational, reflecting the communication of ESG content and its observable correlates rather than verified ESG performance or a demonstrated causal signalling mechanism. Within these boundaries, the study establishes a reproducible basis for evaluating how firms frame ESG at market entry and for testing, in future work, whether more balanced, specific, and credible disclosure affects subsequent capital-market and sustainability outcomes.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.sebi.gov.in/sebiweb/home/HomeAction.do?doListing=yes&sid=3&ssid=15&smid=12, NSE Website, and Prowess IQ CMIE Database.

Author contributions

VS: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. MK: Conceptualization, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. Open Access publication funding provided by Vellore Institute of Technology.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Correction note

This article has been corrected with minor changes. These changes do not impact the scientific content of the article.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frsus.2026.1954577/full#supplementary-material

References

Abbas, Y. A., Ahmad-Zaluki, N. A., and Mehmood, W. (2025). Community and environment disclosures and IPO long-run share price performance. *J Financ Report Account* 23, 1001–1023. doi: 10.1108/JFRA-07-2022-0244

Allen, F., and Faulhaber, G. R. (1989). Signalling by underpricing in the IPO market. *J. Financ. Econ.* 23, 303–323. doi: 10.1016/0304-405X(89)90060-3

Alodat, A. Y., and Hao, Y. (2025). Environmental, social and governance (ESG) disclosure and firm performance: moderating role of board gender diversity and sustainability committee. *Sustain. Dev.* 33, 636–651. doi: 10.1002/sd.3126

Antweiler, W., and Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *J. Finance* 59, 1259–1294. doi: 10.1111/j.1540-6261.2004.00662.x

Arena, C., Bozzolan, S., and Michelon, G. (2015). Environmental reporting: transparency to stakeholders or stakeholder manipulation? An analysis of disclosure tone and the role of the board of directors. *Corp. Soc. Responsib. Environ. Manag.* 22, 346–361. doi: 10.1002/csr.1350

Babu, R. C., and Dsouza, A. E. C. (2021). Post listing IPO returns and performance in India: an empirical investigation. *J Financ Stud Res.* 2021, 1–20. doi: 10.5171/2021.418441

Bae, S. M., Masud, M. A. K., and Kim, J. D. (2018). A cross-country investigation of corporate governance and corporate sustainability disclosure: a signaling theory perspective. *Sustainability.* 10, 2611. doi: 10.3390/su10082611

Baier, P., Berninger, M., and Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: a textual analysis. *Financ. Mark. Inst. Instrum.* 29, 93–118. doi: 10.1111/fmii.12132

Beatty, R. P., and Ritter, J. R. (1986). Investment banking, reputation, and the underpricing of initial public offerings. *J. Financ. Econ.* 15, 213–232. doi: 10.1016/0304-405X(86)90055-3

Benveniste, L. M., and Spindt, P. A. (1989). How investment bankers determine the offer price and allocation of new issues. *J. Financ. Econ.* 24, 343–361. doi: 10.1016/0304-405X(89)90051-2

Bernard, S., Singal, M., Dias, A., Ho, J. A., and Zizka, L. (2025). Sustainable horizons: navigating ESG reporting challenges in the hospitality and tourism industry. *J. Hosp. Tour. Res.* 49, 1141–1158. doi: 10.1177/10963480251352232

Bhabra, H. S., and Pettway, R. H. (2003). IPO prospectus information and subsequent performance. *Financ. Rev.* 38, 369–397. doi: 10.1111/1540-6288.00051

Bhattacharya, A., Chakrabarti, B. B., Ghosh, C., and Petrova, M. (2020). Innovations in financing: the impact of anchor investors in Indian IPOs. *Eur. Financ. Manag.* 26, 1059–1106. doi: 10.1111/eufm.12257

Carter, R. B., Dark, F. H., and Singh, A. K. (1998). Underwriter reputation, initial returns, and the long-run performance of IPO stocks. *J. Finance* 53, 285–311. doi: 10.1111/0022-1082.104624

Carter, R., and Manaster, S. (1990). Initial public offerings and underwriter reputation. *J. Finance* 45, 1045–1067. doi: 10.1111/j.1540-6261.1990.tb02426.x

Chhatwani, M. (2021). Readability and sentiment analysis of financial statements: evidence from India. *SCMS J Indian Manag.* 18, 87–94.

Connelly, B. L., Certo, S. T., Ireland, R. D., and Reutzel, C. R. (2011). Signaling theory: a review and assessment. *J. Manag.* 37, 39–67. doi: 10.1177/0149206310388419

Cornelli, F., Goldreich, D., and Ljungqvist, A. (2006). Investor sentiment and pre-IPO markets. *J. Finance* 61, 1187–1216. doi: 10.1111/j.1540-6261.2006.00870.x

Crain, N., Parrino, R., and Srinivasan, R. (2021). Uncertainty, prospectus content, and the pricing of initial public offerings. *J. Empir. Financ.* 64, 160–182. doi: 10.1016/j.jempfin.2021.08.007

Debnath, P., Debi, H., Goel, A., Das, M., Saha, A., Das, S., et al. (2025). Ramification of environmental social and governance performance on financial performance: evidence from emerging market. *J. Gov. Regul.* 14:58. doi: 10.22495/jgrv14i2art6

Del Gesso, C., and Lodhi, R. N. (2025). Theories underlying environmental, social and governance (ESG) disclosure: a systematic review of accounting studies. *J. Account. Lit.* 47, 433–461. doi: 10.1108/JAL-08-2023-0143

Derrien, F. (2005). IPO pricing in "hot" market conditions: who leaves money on the table? *J. Finance* 60, 487–521. doi: 10.1111/j.1540-6261.2005.00736.x

DiMaggio, P., and Powell, W. (1983). *The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields [SSRN Scholarly Paper]*. Rochester, NY: Social Science Research Network.

Fenili, A., and Raimondo, C. (2021). ESG and the pricing of IPOs: does sustainability matter. *SSRN Electron. J.* doi: 10.2139/ssrn.3860138

Ferris, S. P., Hao, G. Q., and Liao, S. M.-Y. (2013). The effect of issuer conservatism on IPO pricing and performance*. *Rev. Finance* 17, 993–1027. doi: 10.1093/rof/rfs018

Friede, G., Busch, T., and Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *J. Sustain. Finance Invest.* 5, 210–233. doi: 10.1080/20430795.2015.1118917

Fu, M., Yu, D., and Zhou, D. (2023). Secret recipe of IPO survival: ESG disclosure and performance. *Financ. Mark. Inst. Instrum.* 32, 3–19. doi: 10.1111/fmii.12169

Gallego-Álvarez, P. I., and Ortas, P. E. (2017). Corporate environmental sustainability reporting in the context of national cultures: a quantile regression approach. *Int. Bus. Rev.* 26, 337–353. doi: 10.1016/j.ibusrev.2016.09.003

Grewal, J., Riedl, E. J., and Serafeim, G. (2019). Market reaction to mandatory nonfinancial disclosure. *Manag. Sci.* 65, 3061–3084. doi: 10.1287/mnsc.2018.3099

Grewal, J., and Serafeim, G. (2020). Research on corporate sustainability: review and directions for future research. *Found Trends® Account* 14, 73–127. doi: 10.1561/1400000061

Guthrie, J., Cuganesan, S., and Ward, L. (2007). "Legitimacy theory: Asia Pacific interdisciplinary research in accounting conference (5th: 2007)," in *Proc Fifth Asia Pac Interdiscip Res Account Conf*, eds. S. Lawrence and M. J. Milne (Auckland, Publisher: APIRA 2007 Organising Committee), 1–35.

Healy, P. M., and Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: a review of the empirical disclosure literature. *J. Account. Econ.* 31, 405–440. doi: 10.1016/S0165-4101(01)00018-0

Heichl, V., and Hirsch, S. (2023). Sustainable fingerprint – using textual analysis to detect how listed EU firms report about ESG topics. *J. Clean. Prod.* 426:138960. doi: 10.1016/j.jclepro.2023.138960

Hummel, K., and Schlick, C. (2016). The relationship between sustainability performance and sustainability disclosure – reconciling voluntary disclosure theory and legitimacy theory. *J. Account. Public Policy* 35, 455–476. doi: 10.1016/j.jaccpubpol.2016.06.001

Kamugisha, D. K., and Sun, H. (2025). Environmental, social, and governance ratings and corporate financial performance: evidence from emerging economies. *Total Qual. Manag. Bus. Excell.* 36, 1480–1498. doi: 10.1080/14783363.2025.2563668

Kang, E., and Lam, N. B. (2023). The impact of environmental disclosure on initial public offering underpricing: sustainable development in Singapore. *Corp. Soc. Responsib. Environ. Manag.* 30, 119–133. doi: 10.1002/csr.2342

Kim, J., and Yang, W. (2026). Environmental, social, and governance (ESG) research: a systematic review of recent trends (2020–2024). *Sustain. Dev.* 34, 1872–1900. doi: 10.1002/sd.70370

Kirubakaran, J., and Joseph, V. S. (2026). Greening the primary market: ESG disclosures and pricing dynamics of Indian IPOs. *Corp. Soc. Responsib. Environ. Manag.* 33, 4055–4069. doi: 10.1002/csr.70372

Kócziás, C., and Demirel, P. (2025). The impact of green skills and green capabilities on firms' financial performance: a systematic literature review. *J. Innov. Knowl.* 10:100846. doi: 10.1016/j.jik.2025.100846

Koseoglu, M. A., Uyar, A., Kilic, M., Kuzey, C., and Karaman, A. S. (2021). Exploring the connections among CSR performance, reporting, and external assurance: evidence from the hospitality and tourism industry. *Int. J. Hosp. Manag.* 94:102819. doi: 10.1016/j.ijhm.2020.102819

Krinitz, J., and Neumann, D. (2021). "Decision analytics for initial public offerings: how filing sentiment influences stock market returns," in *Market Engineering [Internet]*, eds. H. Gimpel, J. Krämer, D. Neumann, J. Pfeiffer, S. Seifert and T. Teubneret al. (Cham: Springer International Publishing), 45–67.

Lee, T., Kang, S., Utami, D. W., and Na, B. (2026). Mapping corporate environment, social, and governance discourses: analysis of Korean companies' sustainability reports (2014–2024). *Corp. Soc. Responsib. Environ. Manag.*:csr.70760. doi: 10.1002/csr.70760

Loughran, T., and McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *J. Financ. Econ.* 109, 307–326. doi: 10.1016/j.jfineco.2013.02.017

Manes-Rossi, F., and Nicolò, G. (2022). Exploring sustainable development goals reporting practices: from symbolic to substantive approaches—evidence from the energy sector. *Corp. Soc. Responsib. Environ. Manag.* 29, 1799–1815. doi: 10.1002/csr.2328

Menon, D. G., Gunasekar, S., and Gopalakrishnan, B. N. (2026). ESG disclosure and financial performance in Indian hospitality firms using text mining. *Discov. Sustain.* doi: 10.1007/s43621-026-03994-0

Michelon, G., Pilonato, S., and Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: an empirical analysis. *Crit. Perspect. Account.* 33, 59–78. doi: 10.1016/j.cpa.2014.10.003

Monteiro, A. P., Pereira, C., and Barbosa, F. M. (2023). Environmental disclosure on mandatory and voluntary reporting of Portuguese listed firms: the role of environmental certification, lucratively and corporate governance. *Meditari Account. Res.* 31, 524–553. doi: 10.1108/MEDAR-09-2020-1001

Mulchandani, K., Mulchandani, K., Iyer, G., and Lonare, A. (2022). Do equity investors care about environment, social and governance (ESG) disclosure performance? Evidence from India. *Glob. Bus. Rev.* 23, 1336–1352. doi: 10.1177/09721509221129910

Nicolò, G., Zampone, G., Sannino, G., and De Iorio, S. (2022). Sustainable corporate governance and non-financial disclosure in Europe: does the gender diversity matter? *J. Appl. Account. Res.* 23, 227–249. doi: 10.1108/JAAR-04-2021-0100

Nirino, N., Santoro, G., Miglietta, N., and Quaglia, R. (2021). Corporate controversies and company's financial performance: exploring the moderating role of ESG practices. *Technol. Forecast. Soc. Change* 162:120341. doi: 10.1016/j.techfore.2020.120341

Oware, K. M., and David Kweku Botchway, K. (2023). Exchange and moral capital of CSR disclosure and financial distress likelihood of family management firms: evidence from India. *Manag. Res. Rev.* 46, 625–646. doi: 10.1108/MRR-09-2021-0694

Pedersen, L. H., Fitzgibbons, S., and Pomorski, L. (2021). Responsible investing: the ESG-efficient frontier. *J. Financ. Econ.* 142, 572–597. doi: 10.1016/j.jfineco.2020.11.001

Reber, B., Gold, A., and Gold, S. (2022). ESG disclosure and idiosyncratic risk in initial public offerings. *J. Bus. Ethics* 179, 867–886. doi: 10.1007/s10551-021-04847-8

Reimers, N, and Gurevych, I Sentence-BERT: sentence Embeddings using Siamese BERT-networks *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* Hong Kong, China Association for Computational Linguistics (2019). p. 3980–3990

Ritter, J. R., and Welch, I. (2002). A review of IPO activity, pricing, and allocations. *J. Finance* 57, 1795–1828. doi: 10.1111/1540-6261.00478

Sahoo, S., and Rajib, P. (2010). After market pricing performance of initial public offerings (IPOs): Indian IPO market 2002–2006. *Vikalpa J. Decis. Mak.* 35, 27–44. doi: 10.1177/0256090920100403

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., and Leippold, M. (2024). Bridging the gap in ESG measurement: using NLP to quantify environmental, social, and governance communication. *Financ. Res. Lett.* 61:104979. doi: 10.1016/j.frl.2024.104979

Sehgal, V., Garg, N., and Singh, J. (2022). Impact of environmental and social reporting and performance on financial performance of a firm: an Indian study. *Paradigm Manag. Res. J.* 26, 99–118. doi: 10.1177/09718907221126429

Seng, J. L., Yang, P. H., and Yang, H. F. (2017). Initial public offering and financial news. *J Inf Telecommun.* 1, 259–272. doi: 10.1080/24751839.2017.1347762

Shmueli, G., and Koppius, O. R. (2011). Predictive analytics in information systems research1. *MIS Q.* 35, 553–572. doi: 10.2307/23042796

Shobhwani, K., and Lodha, S. (2024). Impact of ESG disclosure scores on financial, operating and market-based performance: evidence from NSE-100 companies. *NMIMS Manag. Rev.* 32, 211–223. doi: 10.1177/09711023241306219

Sinha Ray, R., and Goel, S. (2023). Impact of ESG score on financial performance of Indian firms: static and dynamic panel regression analyses. *Appl. Econ.* 55, 1742–1755. doi: 10.1080/00036846.2022.2101611

Spence, M. (1978). Job market signaling. *In Uncertainty in economics*. Academic Press, 281–306.

Suchman, M. C. (1995). Managing legitimacy: strategic and institutional approaches. *Acad. Manag. Rev.* 20:571. doi: 10.2307/258788

Suieubayeva, S., Sánchez-García, E., Martínez-Falcó, J., Marco-Lajara, B., Sadenova, A., and Montalvo-Falcón, J. V. (2025). Green leaders and global change: uncovering the drivers of corporate environmental sustainability. *Environ. Dev.* 54:101148. doi: 10.1016/j.envdev.2025.101148

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: quantifying language to measure firms' fundamentals. *J. Finance.* 63, 1437–1467. doi: 10.1111/j.1540-6261.2008.01362.x

Veeravel, V., Murugesan, V. P., and Narayanamurthy, V. (2024). Does ESG disclosure really influence the firm performance? Evidence from India. *Q. Rev. Econ. Finance* 95, 193–202. doi: 10.1016/j.qref.2024.03.008

Zhang, Q., Ding, R., Chen, D., and Zhang, X. (2023). The effects of mandatory ESG disclosure on price discovery efficiency around the world. *Int. Rev. Financ. Anal.* 89:102811. doi: 10.1016/j.irfa.2023.102811