Space ride or space race?

The impact of industry leader innovation on peer firm value

Yuran Chen
Accounting and Finance Division
University of Manchester
Yuran.chen@postgrad.manchester.ac.uk

Edward Lee
Accounting and Finance Division
University of Manchester
edward.lee@manchester.ac.uk

Vlad Andrei Porumb
Accounting and Finance Division
University of Manchester
vlad-andrei.porumb@manchester.ac.uk

June 2025

We are grateful to Ole-Kristian Hope, Ioannis Ioannou, and Andrew Pierce for helpful comments.

Space ride or space race?

The impact of industry leader innovation on peer firm value.

Abstract

We examine whether and how news relating to innovation in space industry influences peer firm valuation through two competing channels, i.e. *(i)* partial excludability effect and *(ii)* competition effect. Under the partial excludability effect, investors are likely to anticipate that peer firms can benefit from the technological innovation created by third parties and therefore attach valuation premium to peer firms following successful external technological innovations. Based on the competition effect, investors are likely to perceive that peer firms are threatened by third parties, and thus attach valuation discount (premium) to successful (failed) external technological innovations. Our identification strategy draws on exogenous news relating to rocket launches of *SpaceX*, as industry leader, over the 2010-2023 period. Among S&P500 firms with higher R&D expenditures, we provide evidence consistent with the partial excludability effect. The effect is more pronounced both among firms in the aerospace and defense sector, where technology spillover is likely to be stronger, and also for firms with lower analyst coverage, where information spillover is high. Our evidence suggests that accounting disclosures of R&D expenditures can affect investors' interpretation of both technology and information spillovers.

Keywords: innovation, technology, space economy

Introduction

In modern capital markets, innovation has emerged as a critical driver of firm value and economic growth (Simeth & Cincera, 2016; Bloom et al, 2020; Barth & Gee, 2024). This leads investors to increasingly rely on innovation-related information to adjust their expectations about firms' future performance (Kim & Valentine, 2023; Chu et al., 2024). Nonetheless, due to the unique characteristics of innovation, it is not obvious how the information regarding *peer firms'* innovation is priced by the market (Arrow, 1972; Romer, 1990; Aghion & Tirole, 1994). Specifically, innovation information can create economic value and generate positive externalities through knowledge spillovers to industry peers (Dyer et al., 2023; Kim & Valentine, 2021), due to its nonrival and partially excludable nature[1] (Glaeser & Lang, 2024). On the other hand, innovation information can negatively affect peer firms if it signals hard-to-replicate competitive advantages in long term development of technology (Cao et al., 2018). These contrasting effects imply that investor responses depend on whether they interpret the disclosure as a spillover opportunity or a competitive signal. Therefore, identifying which effect dominates in a given setting is essential to understanding how innovation information is priced. However, extant empirical accounting studies focus on the effect of innovation on the disclosing firms (Hopenhayn & Squintani, 2016; Kim & Valentine; 2023; Chu, 2024),[2] offering limited insight into how investors evaluate

---

[1] Nonrivalry refers to the fact that use by one party does not physically diminish the quality or availability of the innovation for others. Conversely, partial excludability refers to the innovation owner's inability to fully limit others' legal access to their innovation. Innovations are partially excludable because the endogenous legal institutions designed to protect and encourage innovation, such as the patent system or trade secret law, offer incomplete protections.

[2] The notable exception is Zhang (2024), who assesses the real effect of peer companies' innovation disclosure. Nonetheless, this study does not explore the manner In which the investors price the peer firm innovation information.

peer firms' innovation. In this study, we fill this gap by making use of the unique setting offered by the developing space industry to examine how information regarding firms' innovation impact peer firms' market outcomes.

The commercial space sector offers a unique opportunity to perform our analysis, as it has experienced rapid growth, with an estimated annual rate of 9% and over $270 billion in private investment since 2015 (Space Capital, 2023). This surge in innovation generates broader market implications that extend well beyond direct competitors, influencing firms across a wide range of sectors. The commercial space sector *(i)* provides transparency around both innovation successes and failures and *(ii)* allows the clear identification of the industry leader in terms of technological innovation. Regarding the former, all space launches in the U.S. are regulated by the Federal Aviation Administration (FAA), which requires licensed commercial space operators to submit an FAA-approved mishap plan.[3] This regulatory structure ensures that failure information is transparent and accessible to both public and industry peers, making it an ideal setting for studying peer effects. Regarding the latter, among commercial space operators, *SpaceX* is the clear dominant player[4] (McDowell, 2020). As the industry leader, *SpaceX* attracts significant investor and media attention, suggesting that its successes and failures are closely followed by competitors and their stakeholders. This visibility makes *SpaceX* a compelling event study object for examining how market participants interpret peer firm innovation outcomes.

[3] Following any failure, the operator must file a written report to the FAA within five days, and the FAA publicly discloses key incident details.
[4] Out of 855 licensed space launches approved by FAA, 479 are conducted by SpaceX, followed by Rocket Lab's 62 licensed launches.

Current research on peer firms' innovation disclosures is limited in that is has so far focused mostly on the impact of successful peer innovation. For example, Zhang (2024) examines clinical trial disclosures in the pharmaceutical industry and finds that voluntary disclosure of peer firms' preclinical success influences the R&D decisions of firms within the same therapeutic area. Similarly, studies using patent data often rely on disclosed, successful innovations that result in patent filings. The limited existing research on the effects of peer firm innovation failure primarily examines the termination of related projects by focal firms (Krieger, 2021; Xiong et al., 2024). This means that the role of peer firm failures on capital market remains underexplored, in part because such events are seldom disclosed and often receive limited investor attention compared to successful innovations. We therefore differentiate from the extant studies by assessing the market reaction around space-related innovation success and failures of peer firms, not the focal firms' subsequent operating choices.

As commercial space operations become increasingly material to publicly listed firms and capital markets, there is a growing need for quantitative research that investigates how space news are priced by investors (Tucker & Alewine, 2024). The current studies on space accounting mainly falls in the qualitative area, examining the role of accounting in the space sector (Alewine, 2020; Tucker & Alewine, 2023; Tucker & Alewine, 2024), focusing on accountability (Damjanov, 2018; Tucker & Alewine, 2022; Lee & Fong, 2024; Modell, 2024) and performance management (Tucker & Alewine, 2023; Florio et al, 2024). While these studies offer important insights into the conceptual understanding of space accounting, the economic consequences of space activities remain underexplored. This study addresses the gap by empirically assessing whether space-related news generate systematic market reactions and examining how firm-level and industry-level characteristics moderate the market effect.

In this study, we seek to examine how investors respond to industry leader innovation news by disentangling two competing channels: the partial excludability effect and the competition effect. According to the intra-industry information transfer theory (Foster, 1981), information disclosed by one firm can affect the valuation of peer firms when investors perceive economic linkages across firms. However, the direction and intensity of this effect depend on how investors interpret the implications of innovation. The partial excludability perspective suggests that innovation benefits will extend beyond the disclosing firm when peer firms have the capability to absorb and apply external knowledge (Arrow, 1972; Cohen & Levinthal, 1990). In contrast, the competition perspective posits that innovation by an industry leader will be viewed as a strategic threat that reshapes relative positioning within the industry (Leary & Roberts, 2014; Cao et al., 2018). When the former channel dominates, investors are likely to reward peer firms following successful innovation but remain unresponsive to failures.[5] When the latter dominates, investors will react in opposite directions depending on whether the innovation is successful or not. Guided by this conceptual distinction, we develop two hypotheses that reflect the contrasting predictions implied by the partial excludability and competition effects.

[5] Negative news regarding the industry peers' innovation is unlikely to indicate that the future cash-flows of the focal firm are affected.

In order to test how peer innovation news impacts capital market and to examine which channel, partial excludability effect or competition effect, dominates investor interpretation, we hand-collect the major *SpaceX* successful (positive news) and failed (negative news) launches according to Google trend index between 2010 and 2023. We document a positive market reaction of 0.05 percent cumulative abnormal return (CAR) for S&P500 companies with higher R&D expenditure after successful launches, and no significant reaction to failed launches. This finding supports the partial excludability channel, whereby investor interpretate exogenous innovation as knowledge spillover and firms with greater absorptive capacity are perceived as better positioned to benefit from external innovation. The cross-sectional analyses provide further insight into this finding. For firms in the aerospace industry with high R&D expenditure, we observe a CAR increase of 1.7 percent following peer firm success, suggesting that technology spillovers are more pronounced when firms operate in technologically interrelated industries. However, we document a *negative* CAR of 0.3 percent for aerospace firms with low R&D expenditure in response to peer success, indicating that investors recognize peer firm innovation as a competitive threat when the focal firm lacks absorptive capacity. Additionally, we find that the R&D-related return premium is more prominent among firms with lower analyst coverage, indicating that the information spillover function of R&D disclosures plays a more prominent role when firms are less closely monitored by external information intermediaries (Lang & Lundholm, 1996).

This study contributes to the emerging literature on the capital market effect of peer firm innovation disclosure (Tseng, 2022; Glaeser & Lang, 2024). While prior research has primarily examined the direct market consequences of a firm's own innovation disclosure (Chu et al. 2024; Zheng 2025), this paper extends the literature by exploring how market participants respond to innovation from the industry leader. In doing so, this paper introduces a broader framework in which R&D disclosure not only signals internal innovation but also functions as a channel through which firms absorb and respond to external innovation signals from peers.

Formally, we offer three main contributions. First, we answer to the Barth & Gee (2024)'s question of when do parties outside the firm have information to evaluate partial exclusivity. While prior literature has theoretically established that innovation can generate spillovers due to its nonrival and partially excludable nature, this study provides empirical evidence showing that investors assign greater value to peer firm innovation news when they perceive the focal firm to have the capacity to absorb and capitalize on external knowledge. In addition, this study suggests that the R&D disclosure serve a functional role in facilitating the absorption of external knowledge, allowing firms to integrate both technological and informational spillovers into their valuation dynamics. This highlights that the market's valuation of partial excludability is shaped by a firm's absorptive capacity and underscores the central role of R&D disclosure in converting external innovation signals into firm-specific valuation effects.

Second, this study expands the interpretation of R&D disclosure by identifying its' third functional role. While previous research has primarily treated R&D as a signal of mispricing or a proxy for firm-specific risk (Lev & Sougiannis, 1996; Chambers et al., 2002), our findings suggest that R&D disclosure also serves as a receptor of peer firm innovation. Specifically, it enables firms to absorb and incorporate value-relevant external innovation through both technological and informational channels. This perspective broadens the theoretical understanding of how disclosure facilitates the translation of industry-level innovation into firm-level valuation effects.

Third, this study contributes to the emerging literature on space accounting by introducing an empirical, capital market–based approach to understand how space-related information is processed by investors. By doing so, this study moves beyond conceptual and qualitative discussions in the space sector, and opens up new avenues for understanding how financial markets incorporate space innovation into firm valuation.

Institutional background

The space economy, as defined by the Organisation for Economic Co-operation and Development (OECD), refers to a full range of activities and the use of resources that generate value and benefits for humanity through the exploration, research, management, and utilization of space.[6] This sector has experienced tremendous growth in recent years, with its total market value reaching $630 billion in 2023 and projected to surpass $1.8 trillion in the coming decades.[7] A key driver of this expansion is the shift from government-led space initiatives to private-sector participation. The innovations in satellite technology and the declining cost of space launches have made

[6] OECD. (n.d.). Space Economy. Retrieved from https://www.oecd.org/en/topics/space-economy.html
[7] McKinsey & Company. (2023). Space: The $1.8 trillion opportunity for global economic growth. Retrieved from https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/space-the-1-point-8-trillion-dollar-opportunity-for-global-economic-growth

access to space more feasible for private enterprises (Adilov et al., 2022). Companies like SpaceX and Blue Origin gradually take on roles traditionally held by national space agencies such as NASA and the European Space Agency (ESA), accelerating commercialization and investment in the industry.

In recent years, an increasing number of S&P 500 firms have become directly or indirectly involved in the space economy. According to OECD (2019), sectors such as telecommunications, logistics, and agriculture increasingly depend on satellite technology for core operations. As a result, commercial space launches can generate meaningful spillover effects across a broad set of publicly listed firms. A successful launch would signal technological advancement and reliability, benefitting firms with current or planned satellite deployments. Conversely, a failed launch can raise concerns about project delays and service interruptions[8]. This widespread exposure underscores the relevance of space launch events not only for aerospace firms but also for the broader market participants.

Literature Review

Peer Innovation Effect

This study is closely related to the emerging literature on the effect of peer firm innovation. Due to the nonrival and partially excludable nature of innovation, firms are

[8] For example, on September 1, 2016, a SpaceX Falcon 9 rocket exploded on the launch pad during a routine pre-launch static fire test at Cape Canaveral. The explosion destroyed the AMOS-6 communications satellite, which was built by Israel Aerospace Industries and operated by Spacecom. Facebook had leased capacity on AMOS-6 to provide internet connectivity to underserved regions in sub-Saharan Africa as part of its Internet.org initiative. The loss of the satellite not only delayed Facebook's plans to expand internet access but also had significant financial implications for Spacecom, including a drop in its stock price.

often unable to fully internalize the benefits of their technological advancements, thereby generating externalities that would extend to peer firms (Arrow, 1972; Romer, 1990; Glaeser & Lang, 2024). Studies have shown that peer innovation information can affect focal firm's real decisions such as project initiation or termination (Krieger, 2021; Hsu et al., 2022; Zhang, 2024; Xiong et al., 2024). Notably, Zhang (2024) and Xiong et al. (2024) highlights the strategic positioning of the disclosing firm plays a critical role in shaping peer responses, where small firms tend to withdraw from markets in response to disclosures by large firms and large firms would enter markets following disclosures by smaller peers. In parallel, a substantial literature on technological spillovers has documented their broader influence on productivity, firm investment, and asset pricing (Jaffe & Trajtenberg, 1998; Nadiri, 1993; Bloom et al., 2013; Tseng, 2022). These spillovers not only bring technological opportunities but also introduce strategic risks. For instance, peer innovation activity will disincentivize voluntary disclosure when firms view transparency as exposing their competitive vulnerabilities to rivals (Cao et al., 2018; Kao, 2024).

This study differs from Zhang (2024) along two dimensions. First, the outcome of interest varies. Zhang (2024) finds that firms react differently to peer innovation disclosures depending on whether the disclosing firm is a strong or weak competitor within the same market. In contrast, this study examines the boarder capital market effect of innovation news and does not restrict the focal firms to the same industry or direct market competitors. The sample includes firms across a range of sectors beyond aerospace. Second, the nature of the disclosure differs. While Zhang (2024) focuses on peer disclosures related to interim innovation successes, this study investigates the market impact of both positive and negative exogenous innovation events, thereby providing a comprehensive view of how innovation disclosures are interpreted by investors.

Space Accounting

Space accounting is an emerging field that examines how accounting practices intersect with the rapidly developed space economy (Alewine, 2020). Current research in this area has primarily developed through qualitative methodologies and extend established accounting theories to the unique institutional contexts of space (Tucker & Alewine, 2024). Much of this literature focuses on how accounting is used to construct accountability, legitimacy, and governance frameworks in space-related initiatives, particularly in settings marked by high uncertainty and limited regulatory monitoring (Damjanov, 2018; Modell, 2024; Lee & Fong, 2024). These studies contribute valuable conceptual insights into how accounting is mobilized in space sector (Raswant et al., 2025). However, empirical research in this domain remains limited. A small number of studies have begun to explore space accounting through data-driven approaches. For example, Xu et al. (2024) examine the relationship between corporate social responsibility (CSR) practices and R&D quality within the space sector, while Smith & King (2024) analyze media narratives to identify the strategies employed by space agencies to attract private investment in commercial space travel. Despite these contributions, there remains a research gap in understanding how space-related news influence firm-level financial outcomes and investor decision-making, pointing to the need for archival research that evaluates the economic consequences of accounting practices in the space economy.

Hypothesis Development

The intra-industry information transfer theory suggests that information disclosed by one firm can influence the valuation of its industry peers when investors perceive economic linkages across firms (Foster, 1981). This theoretical framework suggests that external innovation information can serve as relevant signals for peer firms' investor, particularly when innovation outcomes are not firm-specific but reflect broader industry trends. Investor reactions to such signals can follow two distinct interpretations: one driven by potential knowledge spillovers, which reflects the partial excludability effect, and the other by strategic rivalry, which reflects the competition effect.

The first perspective posits that the benefits of innovation can extend beyond the innovating firm due to the nonrivalry and partial excludability nature of innovation (Arrow, 1972; Romer, 1990; Glaeser & Lang, 2024). However, according to the absorptive capacity framework, whether peer firms can realize these benefits depends on their absorptive capacity (Cohen & Levinthal, 1990). As a result, a firm's R&D investment plays a key role as its disclosure provides investors with a visible signal of focal firm's absorptive capacity. In this sense, accounting disclosure of R&D expenditure acts as a bridge between external innovation and firm-specific valuation, channeling the knowledge spillover into market perceptions.

An alternative interpretation is offered by the competition effect, which suggests that innovation by an industry leader is not seen as a shared opportunity but rather as a competitive threat to peer firms. If the innovation breakthrough is perceived to strengthen the industry leader's advantage, it would reduce expected future cash flows for its competitors and thus lead to a negative market reaction for the peer firms (Leary & Roberts, 2014; Grennan, 2019). Conversely, an innovation failure by the industry leader would be interpreted as a weakening of their advantage, thereby easing competitive pressure on focal firms (Cao et al., 2018). This logic suggests that both favorable and unfavorable innovation news from industry leaders matter for investors' valuation of focal firms, and that the direction of the market response will differ depending on whether the event signals strength or weakness.

The competing channels within the frameworks of partial excludability and competition effects lead to a divergence in empirical predictions. If the market response reflects partial excludability, then investor reactions should be concentrated around successful innovation only, as firms with sufficient absorptive capacity are expected to benefit from positive spillovers. In contrast, innovation failure would not materially affect the valuation of peers, since unsuccessful knowledge does not transfer or affect future cash flows. However, if investor responses are shaped by competitive concerns, then both types of innovation news become consequential. In this case, favorable news is likely to be interpreted as a threat to peer firms' relative position, while unfavorable news reduces perceived competitive pressure. As a result, the direction of the market reaction would be opposite for success and failure. This theoretical tension motivates the following two hypotheses:

H1: *If the partial excludability effect dominates over the competition effect, market reactions among peer firms to industry leader innovation news are likely to be positive for favorable news and insignificant to unfavorable news.*

H2: *If the competition effect dominates over the partial excludability effect, market reactions among peer firms to industry leader innovation news are likely to be in opposite in direction for both favorable and unfavorable news.*

Identifying which effect dominates is essential for understanding how capital markets respond to external innovation. If investors primarily reward firms for their ability to absorb innovation, this highlights the value of developing and disclosing absorptive capacity through R&D investments. In contrast, if competitive threats drive investor reactions, it suggests that innovation by peers may reduce focal firm value, particularly for those lacking technological strength. Distinguishing between these two mechanisms helps clarify how innovation spillovers are priced in financial markets and therefore provides insights on how firm-level innovation strategy and the role of R&D disclosure shapes investor expectations.

Research Design

To examine how peer firm innovation news affects stock market reactions, we employ an event study framework combined with an OLS regression analysis (Brown & Warner, 1985; Borusyak, et al. 2024). The regression model is estimated as follows:

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 Controls_i + \epsilon \quad (1)$$

Where *i* represents firms and *j* represents *SpaceX* launch events. The dependent variable $CAR_{i,j}$ is defined as the cumulative abnormal return over an [0,+1] window of firm *i* around launch event *j*, which is calculated using the adjusted market model with a 200-day estimation window prior to each innovation event (Campbell et al., 1997; Kothari & Warner, 2004). Events are defined as major launch outcomes by *SpaceX* between 2010 and 2023, manually collected from public sources[9]. The detailed list of events is presented in Appendix 1.

The key explanatory variable RD_i captures whether the focal firm's disclose higher R&D expenditure, which measures as the difference between the focal firm's R&D expenditure scaled by total assets and the industry median R&D expenditure in the same year. This variable captures whether a firm is more or less R&D-intensive relative to its peers, which proxies for its absorptive capacity to engage with external innovation.

A relevant set of control variables are further included, represents by $Controls_i$. Specifically, we use three sets of controls, following prior work by Boehmer et al. (1991) and Campbell et al. (2010). The first set of control variable captures firm characteristics that can influence investor responses to peer innovation. Firm size (*SIZE*), leverage (*LEV*), return on assets (*ROA*), capital intensity (*CapInt*), and market-to-book ratio (*MTB*) are included, as previous studies suggest that larger, more profitable, and less financially constrained firms tend to attract greater market attention and will exhibit

[9] The events are selected based on Google Trends Index data.

more stable abnormal returns around external events (Atiase, 1985; Daniel & Titman, 2006; Hennessy & Whited, 2007). In addition, space disclosure (SD), which is calculated as the natural logarithm of 1 plus the number of space-related words in the firm's 10-K report, divided by the total number of words, is also included to mitigate the effect of firm-specific exposure related to the space economy. The second set of control variables relates to the firm's information environment and corporate governance. Prior research suggests that when a firm consistently meets or beats earnings expectations (*MBE*), the market adjusts more conservatively to new information (Skinner & Sloan, 2002). In addition, analyst following (*Analysis*) reflects the degree of external information coverage and visibility, which shapes the accuracy of investor expectations and reduce the element of surprise in response to peer innovation (Lang & Lundholm, 1996; Hope, 2003). Moreover, board characteristics can affect a firm's strategic responsiveness and oversight quality. Firms with a higher proportion of independent directors (*BIND*) and longer board duration (*BDUR*) experience variation in their ability of adapting innovation signals (Coles et al., 2008). The third set of control refer to macroeconomic indicator, which are the U.S. gross domestic product (*GDP*), inflation rates (*Inflation*), interest rates (*Interest*), and the market risk premium (MKT_p). These variables capture broader economic conditions and investor sentiment that would influence stock price (Chen, Roll, & Ross, 1986). The detailed representations of variable definitions are listed in Appendix 2.

To further investigate whether the nature of innovation events moderates the market response, we examine the difference between successful and failed events. A binary variable $FAIL_j$ is introduced, which is set to 1 if the event is a failed innovation (unsuccessful SpaceX rocket launch or unsuccessful test launch) and set to 0 otherwise. This indicator is then interacted with the independent variable RD_i, to test whether investor interpretation of peer innovation differs by event outcome. The extended model is specified as:

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 FAIL_j + \beta_3 RD_i \times FAIL_j + \beta_4 Controls_i + \epsilon \quad (2)$$

Data

We start with an initial sample comprising of all U.S. firms listed in the S&P 500 index from 2010 to 2024. The sample selection process begins by excluding firm-year observations with missing disclosure on R&D expenditure data, as this represents a key explanatory variable in the analysis. Additionally, all observations from the year 2024 are excluded due to incomplete data availability, to avoid introducing temporal imbalance. This selection process generates a sample of 6,909 firm-year observations, as listed in Table 1. The pairwise correlation of all variables is shown in Table 2. The generally low correlations among pairs of variables suggests that multicollinearity is unlikely to bias the regression estimates.

[Insert Table 1 here]

[Insert Table 2 here]

We further make use of multiple data sources to construct our variables. The firm fundamental data is obtained from the *Compustat*, *Center for Research in Security Prices (CRSP)*, and the *Institutional Brokers' Estimate System (I/B/E/S)* databases. The corporate governance data are sourced from the *Institutional Shareholder Services (ISS)* database, while the macroeconomic indicators are retrieved from the *Federal Reserve Economic Data (FRED)* platform maintained by the U.S. Federal Reserve Bank of St. Louis.

In order to capture how the market reacts to industry innovation news, we hand-collect a series of exogenous space events related to *SpaceX*. The initial event set was built using publicly accessible space-related websites to identify journal articles and reports mentioning *SpaceX*'s achievements since 2010, resulting in a preliminary list of 438 space-related events. To refine the selection and identify events that received significant public attention, each event date was cross-referenced with Google Trends data. Events with a Google Trends interest index below 50 were excluded, as they were regarded as having insufficient public recognition, forming a final list of 19 events with 12 success events and 9 failure events.

Empirical Findings

Table 3 provides the results of the baseline regression examining the effect of peer firm innovation news on stock market reactions for S&P500 firms with high vs. low R&D expenditure. Columns (1) to (4) show the sequence of regression model to include three sets of control variables separately. The coefficient of our variable of interest, RD_i, remains consistent and significant across all models. Therefore, these results demonstrate that the positive effect of R&D disclosure on market response to peer innovation is consistent and statistically significant, even after controlling for firm characteristics, governance structure, and macroeconomic factors. Column (4) includes the full set of controls and shows that the coefficient of RD_i is positive and significant at the 10 percent level with magnitude of 0.056, suggesting that a unit increase in R&D expenditure would result an 0.056 percent increase in CAR. Overall, the results in Table 3 suggest that the positive market reaction to industry leader innovation is concentrated among firms with high R&D expenditure. This finding is consistent with the idea that firms with greater innovation investment possess stronger absorptive capacity, allowing them to benefit more from external technological signals.[10]

[Insert Table 3 here]

Table 4 provides the results of the regression examining how the stock market reacts to failed innovation events (*SpaceX* launch failures) among S&P 500 firms with high R&D expenditure. Column (1) to Column (4) presents the baseline regression without control variables to the inclusion of full sets of control variables. In Column (4), the coefficient of RD_i and $FAIL_j$ remains positive and statistically significant at the 10 percent and 5 percent level separately, while the coefficient of interaction term $RD_i \times FAIL_j$ remains significantly negative at the 10 percent level and have similar economic magnitude than the positive effect of RD_i. The Table 4 results suggest that while high-R&D firms are generally rewarded in the market, this advantage is offset when the innovation signal is unfavorable. The similar economic magnitudes of the

[10] Consistent with previous studies (Coles et al., 2008; Cai & Sevilir, 2012), we find that firms with more independent boards tend to experience weaker market reactions to peer innovation news. Furthermore, we find that a higher market risk premium is associated with lower CARs, suggesting that during periods of elevated market-wide expected returns, investors respond less positively to firm-specific peer innovation signals. This is consistent with the idea that macro-level risk sentiment can attenuate attention to idiosyncratic news (Chen et al., 1986; Nguyen & Puri, 2021).

positive coefficient on RD_i and the negative coefficient on the interaction term $RD_i \times FAIL_j$ indicate that the market exhibits a muted response to industry leader failure. This absence of significant reaction implies that unfavorable innovation events lack transferable knowledge spillovers through partial excludability effect, limiting investors' ability to reassess peer firm value.

[Insert Table 4 here]

Collectively, the baseline results show robust results across all specification, providing strong support for Hypothesis 1, where Table 3 results suggest a positive market reaction among peer firms to industry leader innovation for favorable news and Table 4 results suggest an insignificant market reaction to unfavorable news. Taken together, the results lend support to the interpretation that peer firm market reactions are primarily driven by the partial excludability effect rather than by competitive concerns.

Further Analysis

Building on the baseline estimation, we conduct a series of cross-sectional analyses to investigate heterogeneity in investor responses to peer firm innovation.

Using the subsample of aerospace industry, we examine whether the relationship between disclosure of R&D expenditure and stock price reactions varies by industry proximity to the innovating firm. Specifically, we focus on the subsample of S&P 500 firms that are in the aerospace industry. Firms in this sector are closely linked to each other through shared technologies such as engines, satellite platforms, and suppliers

(OECD, 2020). This high interdependence makes innovation by one firm more likely to generate technology spillover to others in the industry and makes innovation news in this sector more relevant to investors while evaluating firms (Kapoor & Lee, 2013). As a result, peer innovation in aerospace sector is not only seen as an isolated success but often signals changes to the industry's competitive and technological landscape.

Due to regulatory constraints and long development cycles, innovation in the aerospace industry tends to have a slow-moving and cumulative nature (Bonvillian & Weiss, 2015). In such environment, firms are often locked into specific technological paths, making adaptability become difficult (Arthur, 1989). As a result, firms that lack strong innovation capabilities often struggle to respond when technological shifts occur and do not have the ability to absorb external knowledge spillover (Adner & Kapoor, 2010). By contrast, firms with higher R&D expenditure, suggesting a higher absorption ability, would in better position of tracking external developments and incorporate this innovation into their own operations (Teece, 2007). These difference in absorption capacity would shape how aerospace industry investors interpret *SpaceX's* innovation. Specifically, a signal from *SpaceX* will be viewed as an opportunity for aerospace firms perceived as adaptive, but as a risk for those seen as lagging (Cohen & Levinthal, 1990).

To empirically test our contentions, we define AD_i as an indicator variable equal to 1 if the firm belongs to the aerospace industry. We them interact AD_i with the variable of interest and re-estimate both Equations (1) and (2), forming Equations (3) and (4) as follows:

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 AD_i + \beta_3 RD_i \times AD_i + \beta_4 Controls_i + \epsilon \tag{3}$$

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 FAIL_j + \beta_3 RD_i \times FAIL_j + \beta_4 AD_i + \beta_5 FAIL_j \times AD_i + \beta_6 RD_i \times FAIL_j \times AD_i + \beta_7 Controls_i + \epsilon \quad (4)$$

The regression result of Equation (3) is showed in Table 5. Similar to our previous approach, we sequentially include control groups in our models and the regression results are shown in Columns (1) to (4). In Column (4) we present the full model with firm-level, governance, and macroeconomic controls. The coefficient of the interaction term $RD_i \times AD_i$ remains consistently statistically significant at the 1 percent level, with a magnitude of 1.746. This indicates that within the aerospace sector, firms with higher R&D expenditure benefit more from peer innovation news. This finding provides additional support for Hypothesis 1 by showing that the partial excludability effect is more pronounced when the technological spillover is high. The coefficient of aerospace industry indicator AD_i is negative and significant at the 5 percent level, indicating that aerospace firms experience negative abnormal returns following peer innovation events. This suggests that without sufficient absorptive capacity, aerospace firms are perceived as less equipped to capitalize on external innovation, leading to weaker investor confidence in their competitive positioning. These results collectively show that within the aerospace industry, firms with higher R&D expenditure are rewarded by the market for their potential to leverage peer innovation, while those with weaker innovation capacity are penalized due to increased exposure to competitive pressure.

[Insert Table 5 here]

Table 6 reports the results of the Equation (4), evaluating the effect of failed peer innovation on aerospace firms with high R&D expenditure. The results presented in Columns (1) to (4) are consistent and reinforce the baseline findings. The regression results in Column (4) show that the coefficient of $RD_i \times AD_i$ remains positive and significant at the 1 percent level with magnitude 3.124, while the three-way interaction term $RD_i \times AD_i \times FAIL_j$ is negative and also statistically significant at the 1 percent level with similar magnitude of 3.089. This indicates that although aerospace firms with high R&D expenditure generally benefit from peer innovation, when peer innovation fails this positive market reaction disappears. This finding further confirm Hypothesis 1 where the market reactions among peer firm to industry leader innovation are insignificant to unfavorable new, which indicates the domination of partial excludability effect.

[Insert Table 6 here]

Collectively, the cross-sectional test based on the aerospace industry shows that the effect of excludability effect is more pronounced among firms in the aerospace sector where technology spillover is stronger.

We further focus on the role of the information environment, as proxied by analyst following, in shaping market reactions to peer innovation. Investor responses to peer firm innovation are shaped not only by the focal firm's innovation capacity, but also by the surrounding information environment (Bushee & Noe, 2000). According to the investor recognition hypothesis, firms that are more visible and better followed tend to attract investors with more accurate expectations (Merton, 1987). Analyst following contributes to this process by increasing the visibility of the firm and reducing uncertainty around its key characteristics, including innovation efforts (Lang & Lundholm, 1996; Hope, 2003). In the context of peer innovation news, greater visibility would limit the extent to which investors revise their expectations. For firms with high R&D expenditure, higher analyst coverage indicates that investors are already well informed about their innovation capacity, leading the magnitude of information spillover decrease. As such, when a peer firm experienced major innovation, the market would not revise its expectations about the focal firm upward, because its innovation capability is already priced in (Chen et al., 2002). In some cases, the absence of surprise would even result in relatively muted or negative market reactions, particularly when the focal firm is not directly involved in the breakthrough (Drake et al., 2012). As a result, analyst following can reduce the scope for positive spillover effects from peer innovation, and would lead to more conservative investor responses when expectations are already high.

In order to test the modification effect of information spillover on investor response, we introduce $Analysis_h$, which is a dummy variable equals to 1 if the focal firm has higher analysis following than industry medium and 0 otherwise. We them interact $Analysis_h$ with the variable of interest and re-estimate both Equations (1) and (2), forming Equations (5) and (6) as follows:

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 Analysis_h + \beta_3 RD_i \times Analysis_h + \beta_4 Controls_i + \epsilon \quad (5)$$

$$CAR_{i,j} = \beta_0 + \beta_1 RD_i + \beta_2 FAIL_j + \beta_3 RD_i \times FAIL_j + \beta_4 Analysis_h + \beta_5 FAIL_j \times Analysis_h + \beta_6 RD_i \times FAIL_j \times Analysis_h + \beta_7 Controls_i + \epsilon \quad (6)$$

The regression results of Equation (5) are reported in Table 7, which examines how analyst following moderates the relationship between R&D expenditure disclosure and stock market reactions to peer innovation events. The regression result with different controls are consist. The full model regression result is shown in Column (4), where coefficient of the interaction term $RD_i \times Analysis_h$ is negative and statistically significant at the 5 percent level, suggesting that the positive effect of R&D expenditure is weaker when analyst following is higher. The coefficient of RD_i is positive and statistically significant at the 5 percent level, indicating that high R&D firms experience stronger stock price responses to peer innovation. Collectively, these results indicates that when a firm's innovation capacity is already well known and priced in due to high analyst coverage, the marginal value of peer innovation news diminishes. In this context, analyst following reduces the surprise element in innovation spillovers, thereby attenuating the market's reaction to external innovation events.

[Insert Table 7 here]

Table 8 presents the regression results of Equation (6), which further explores how analyst following affects investor responses to failed peer innovation events among firms with different levels of R&D expenditure. The regression results are consist while including different sets of control. In Column (4), which represents the regression model with full sets of control, the coefficient on RD_i remains positive and significant at the 5 percent level, while the coefficient on the failure indicator $FAIL_j$ is also positive and statistically significant. However, the coefficient on the interaction term $RD_i \times FAIL_j$ is negative and statistically significant at the 5 percent level, suggesting that the beneficial effect of R&D is reduced in the presence of peer innovation failure. Importantly, the three-way interaction term $RD_i \times FAIL_j \times Analysis_h$ is significantly positive at the 5 percent level, indicating that firms with higher R&D and greater analyst following experience less negative reactions when peer innovation fails. This result provides further nuance to the information spillover effect, showing that while analyst following will limit the upside from peer innovation success, it can also help to mitigate the downside risks in the face of peer failure by clarifying firm-specific innovation strength.

[Insert Table 8 here]

Conclusion

This study examines how capital markets interpret space-related innovation news, using *SpaceX* launch events as a novel and externally observable information shock. We contribute to the growing literature on the impact of peer firm innovation by drawing on the theoretical foundations of intra-industry information transfer and the unique characteristic of innovation. This study documents that investors assign greater value to firms perceived as capable of absorbing and responding to peer innovation, consistent with the partial excludability effect. Building on this insight, this study shows that R&D disclosure plays a strategic role as a receptor of peer innovation, enabling firms to convert external breakthroughs into internal value through technological and informational spillovers, thereby extending the existing view that R&D primarily reflects mispricing or risk premium.

The findings illustrate that investor responses to peer innovation are shaped not only by the nature of the innovation itself but also by firm-specific characteristics and industry context. By embedding these dynamics within the context of the space economy, this study not only advances research related to peer firm’s innovation effects but also extends the scope of space accounting into empirical capital market research.

References

Adilov, N., Alexander, P., Cunningham, B., & Albertson, N. (2022). An analysis of launch cost reductions for low earth orbit satellites. *Economics Bulletin*, 42(3), 1561-1574.

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, 31(3), 306–333.

Aghion, P., & Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics*, *109*(4), 1185-1209.

Alewine, H. C. (2020). Space accounting. *Accounting, Auditing & Accountability Journal,* 33(5), 991-1018.

Arrow, K. J. (1972). *Economic welfare and the allocation of resources for invention* (pp. 219-236). Macmillan Education UK.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394), 116–131.

Atiase, R. K. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, 23(1), 21–36.

Barth, M. E., & Gee, K. H. (2024). Accounting and innovation: Paths forward for research. *Journal of Accounting and Economics*, *78*(2-3), 101733.

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find?. *American Economic Review*, *110*(4), 1104-1144.

Boehmer, E., Masumeci, J., & Poulsen, A. B. (1991). Event-study methodology under conditions of event-induced variance. *Journal of Financial Economics*, 30(2), 253-272.

Bonvillian, W. B., & Weiss, C. (2015). Technological Innovation in Legacy Sectors. *Oxford University Press*.

Bonvillian, W. B., & Weiss, C. (2015). Technological Innovation in Legacy Sectors. *Oxford University Press*.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, 91(6), 3253-3285.

Breuer, M., Leuz, C., & Vanhaverbeke, S. (2025). Reporting regulation and corporate innovation. *Journal of Accounting and Economics*, 101769.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(1), 3-31.

Bushee, B. J., & Noe, C. F. (2000). Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research*, 38(Supplement), 171–202.

Cai, Y., & Sevilir, M. (2012). Board connections and M&A transactions. *Journal of Financial Economics*, 103(2), 327–349.

Campbell, C. J., Cowan, A. R., & Salotti, V. (2010). Multi-country event-study methods. *Journal of Banking & Finance*, 34(12), 3078-3090.

Cao, J., Guo, B., & Ritter, J. R. (2018). Information spillovers in equity crowdfunding. *Review of Financial Studies*, 31(10), 3997–4034.

Cao, S. S., Ma, G., Tucker, J. W., & Wan, C. (2018). Technological peer pressure and product disclosure. *The Accounting Review*, 93(6), 95-126.

Chambers, D., Jennings, R. & Thompson, R. (2002). Excess returns to R&D-intensive firms. *Review of Accounting Studies*, 7(2–3): 133–158.

Chen, N. F., Roll, R., & Ross, S. A. (1986). Economic forces and the stock market. *Journal of Business*, 59(3), 383–403.

Chen, S., DeFond, M. L., & Park, C. W. (2002). Voluntary disclosure and the cost of capital: Evidence from management earnings forecasts. *Journal of Accounting Research*, 40(3), 895–921.

Chu, J., He, Y., Hui, K. W., & Lehavy, R. (2024). New product announcements, innovation disclosure, and future firm performance. *Review of Accounting Studies*, 1-32.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128-152.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35(1), 128–152.

Coles, J. L., Daniel, N. D., & Naveen, L. (2008). Boards: Does one size fit all? *Journal of Financial Economic*s, 87(2), 329–356.

Correa, J. A. (2012). Innovation and competition: An unstable relationship. *Journal of Applied Econometrics*, 27(1), 160-166.

Damjanov, K. (2018). Accounting for non-humans in space exploration. *Space Policy*, 43, 18-23.

Dicks, D. L., & Fulghieri, P. (2019). Uncertainty aversion and systemic risk. *Journal of Political Economy*, *127*(3), 1118-1155.

Dicks, D., & Fulghieri, P. (2021). Uncertainty, investor sentiment, and innovation. *The Review of Financial Studies*, 34(3), 1236-1279.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research*, 50(4), 1001–1040.

Dyer, T. A., Glaeser, S., Lang, M. H., & Sprecher, C. (2024). The effect of patent disclosure quality on innovation. *Journal of Accounting and Economics*, *77*(2-3), 101647.

Florio, M., Castelnovo, P., Landoni, M., Lupi, V., Morretta, V., Vurchio, D., ... & Piermaria, M. (2024). Accounting for the impact of space policies: The case of the Italian Space Agency. *Space Policy*, 101597.

Foster, G. (1981). Intra-industry information transfers associated with earnings releases. *Journal of Accounting and Economics*, *3*(3), 201-232.

Glaeser, S. A., & Landsman, W. R. (2021). Deterrent disclosure. *The Accounting Review*, *96*(5), 291-315.

Glaeser, S., & Lang, M. (2024). Measuring innovation and navigating its unique information issues: A review of the accounting literature on innovation. *Journal of Accounting and Economics*, 101720.

Grennan, J. (2019). Dividend payments as a response to peer influence. *Journal of Financial Economics*, 131(3), 549–570.

Hegde, D., & Luo, H. (2018). Patent publication and the market for ideas. *Management Science*, *64*(2), 652-672.

Hope, O.-K. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research*, 41(2), 235–272.

Hopenhayn, H. A., & Squintani, F. (2016). Patent rights and innovation disclosure. *The Review of Economic Studies*, *83*(1), 199-230.

Hughes, J. S., & Pae, S. (2015). Discretionary disclosure, spillovers, and competition. *Review of Accounting Studies*, *20*, 319-342.

Kapoor, R., & Lee, J. M. (2013). Coordinating and competing in ecosystems: How organizational forms shape new technology investments. *Strategic Management Journal*, 34(3), 274–296.

Kim, J., & Valentine, K. (2021). The innovation consequences of mandatory patent disclosures. *Journal of Accounting and Economics*, *71*(2-3), 101381.

Kim, J., & Valentine, K. (2023). Public firm disclosures and the market for innovation. *Journal of Accounting and Economics*, 76(1), 101577.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics*, *132*(2), 665-712.

Lang, M. H., & Lundholm, R. J. (1996). Corporate disclosure policy and analyst behavior. Accounting review, 467-492. Truong, P. (2023). Peer effects and disclosure timing: Evidence from earnings announcements. *The Accounting Review*, 98(3), 427-458.

Lang, M., & Lundholm, R. (1996). Corporate disclosure policy and analyst behavior. *The Accounting Review*, 71(4), 467–492.

Leary, M. T., & Roberts, M. R. (2014). Do peer firms affect corporate financial policy? *Journal of Finance*, 69(1), 139–178.

Lee, Y., & Fong, E. A. (2024). The art of living together: space mining ecosystem, sustainability and accountability. *Accounting, Auditing & Accountability Journal.*

Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, *54*(2), 525-622.

Lev, B. & Sougiannis, T. (1996). The capitalization, amortization and value-relevance of R&D. *Journal of Accounting and Economics*, 21(1): 107–138.

Majid, S., Sultana, N., Abid, G., & Elahi, A. R. (2021). The impact of corporate innovation on abnormal stock returns: The moderating role of investor sentiment. *Academy of Strategic Management Journal,* 20, 1-16.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. The *Journal of Finance*, 42(3), 483–510.

Nguyen, M., & Puri, M. (2021). Macroeconomic conditions, market attention, and stock price reactions to firm news. *Review of Financial Studies*, 34(11), 5421–5460.

OECD. (2020). The space economy in figures: How space contributes to the global economy. OECD Publishing.

Plečnik, J. M., Yang, L. L., & Zhang, J. H. (2022). Corporate innovation and future earnings: does early patent disclosure matter?. *Accounting & Finance*, *62*, 2011-2056.

Raswant, A., Nielsen, B. B., & Buckley, P. J. (2025). Space: a new frontier for international business. *Journal of International Business Studies*, 1-22.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5, Part 2), S71-S102.

Saidi, F., & Žaldokas, A. (2021). How does firms' innovation disclosure affect their banking relationships?. *Management Science*, *67*(2), 742-768.

Simeth, M., & Cincera, M. (2016). Corporate science, innovation, and firm value. *Management Science*, 62(7), 1970-1981.

Space Capital. (2023). Space Investment Quarterly Q4 2022.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and micro foundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.

Tucker, B. P., & Alewine, H. C. (2022). Space for accounting and accountability: realising potential management accounting research contributions to the space sector. *Accounting, Auditing & Accountability Journal,* 35(6), 1353-1374

Tucker, B. P., & Alewine, H. C. (2023). Everybody's business to know about space: cross-disciplinarity and the challenges of the new space age. *Space Policy*, 66, 101573.

Tucker, B. P., & Alewine, H. C. (2023). The roles of management control: lessons from the Apollo program. *Contemporary Accounting Research*, 40(2), 1046-1081.

Tucker, B. P., & Alewine, H. C. (2024). Not because it is easy… accounting research enters the space age. *Accounting, Auditing & Accountability Journal*, 37(5), 1245-1257.

Tucker, B. P., & Alewine, H. C. (2024). Solutions looking for problems? How humanities, arts, and social sciences can inform the space sector. *Space Policy*, 67, 101595.

Zhang, Y. (2024). Corporate R&D investments following competitors' voluntary disclosures: Evidence from the drug development process. *Journal of Accounting Research*, *62*(1), 335-373.

Zheng, Y. (2025). Can investors learn from patent documents? Evidence from textual analysis. *Contemporary Accounting Research*.

APPENDIX

Appendix 1. Definition of Variables

Variable	**Definition**
CAR	Cumulative abnormal return, calculated using the adjusted market model of the [0,1] event window around selective *SpaceX* launch events
RD	R&D expenditure, defined as the difference between the focal firm's R&D expenditure divided by total assets and the industry median R&D expenditure divided by total assets.
FAIL	Event type, defined as 1 if the Elon Musk-related event is classified as a failure event, and 0 otherwise
SD	Space Disclosure, calculated as the natural logarithm of 1 plus the number of space-related words in the firm's 10-K report, divided by the total number of words
SIZE	Firm size, calculated as the natural logarithm of total assets
LEV	Leverage, defined as total debt divided by total assets
ROA	Return on assets, defined as net income divided by total assets
CapInt	Capital intensity, defined as the net property, plant and equipment divided by total assets
MTB	Market to Book ratio, defined as market value of equity divided by book value of equity
MBE	Earnings surprise indicator, equal to 1 if the firm meets or beats analysts' forecasts, and 0 otherwise.
Analyst	Analyst following, measured as the number of analyst following
BIND	Board independence, measured as the number of independent directors on the board divided by total board number.
BDUR	Board duration, calculated as the average number of years served by board members.
BOARD	Board size, calculated as the natural logarithm of 1 plus the number of directors
GDP	GDP, measured by the annual average U.S. Gross Domestic product
Inflation	Inflation, measured as the annual average U.S. 10-Year Breakeven Inflation Rate
Interest	Interest rate, measured by the annual average U.S. Federal Funds Effective Rate
MKT_P	Market premium, calculated as the return on the market portfolio minus the risk-free rate, based on the Fama-French three-factor model

Appendix 2. SpaceX Launch and Test Events

Event Date	Type	Description
2010-06-07	*Success*	Falcon 9 completes its inaugural launch from Cape Canaveral, marking SpaceX’s entry into orbital launch capability.
2012-05-22	*Success*	Falcon 9 launches Dragon capsule, achieving the first commercial docking with the ISS.
2012-07-11	*Failure*	Falcon 9 launch experiences engine failure, resulting in the loss of 20 Starlink satellites.
2013-03-01	*Success*	Falcon 9 launches SpaceX's third cargo resupply mission (CRS-2) to the ISS under NASA's CRS contract.
2014-04-21	*Success*	Falcon 9 launches CRS-3 mission carrying supplies and scientific equipment to the ISS.
2014-06-30	*Success*	Falcon 9 launches six Orbcomm OG2 satellites; first flight with reusability test hardware onboard.
2015-06-28	*Failure*	Falcon 9 explodes mid-flight shortly after liftoff during CRS-7 mission, destroying NASA cargo.
2015-12-21	*Success*	Falcon 9 booster successfully lands upright for the first time at Cape Canaveral, advancing reusability.
2016-09-01	*Failure*	Falcon 9 explodes during pre-launch testing for the AMOS-6 satellite mission at Cape Canaveral.
2017-12-26	*Success*	Falcon 9 launches Iridium NEXT satellites from Vandenberg Air Force Base.
2018-01-07	*Failure*	Falcon 9 launches the classified Zuma payload; the satellite is presumed lost after deployment.
2018-02-06	*Success*	Falcon Heavy completes its maiden launch, sending a Tesla Roadster into heliocentric orbit.
2019-04-20	*Failure*	Dragon 2 spacecraft explodes during static fire test at Cape Canaveral Launch Complex 40.
2019-05-02	*Failure*	SpaceX confirms Crew Dragon vehicle was destroyed in April's test fire incident.
2019-08-06	*Success*	Falcon 9 launches Amos-17 communications satellite to geostationary orbit from Cape Canaveral.
2020-05-30	*Success*	Falcon 9 launches Demo-2 mission with two NASA astronauts aboard, marking the first U.S. crewed flight since 2011.
2021-09-16	*Success*	Falcon 9 launches the Inspiration4 mission, placing the first all-civilian crew into Earth orbit.
2022-01-31	*Success*	Falcon 9 launches the COSMO-SkyMed FM2 satellite; same day Elon Musk announces Twitter acquisition.
2023-04-20	*Failure*	Starship completes first integrated test flight but explodes minutes after liftoff.

Table 1

Summary Statistics

	N	Mean	SD	p25	Median	p75
CAR	6909	-0.022	1.672	-0.109	-0.001	0.050
RD	6909	0.152	0.811	-0.084	0.000	0.0076
SD	6909	0.538	1.063	0.000	0.000	0.693
SIZE	6909	0.005	0.982	-.0702	-0.104	0.558
LEV	6881	0.001	1.041	-0.562	-0.057	0.426
ROA	6461	0.029	0.977	-0.514	-0.083	0.493
CapInt	6762	0.018	1.011	-0.749	-0.414	0.556
MTB	6461	0.007	0.960	-0.04	-0.018	9.035
MBE	6722	0.77	0.421	1	1	1
Analysis	6722	18.125	7.709	13	18	23
BIND	5772	0.016	0.975	-0.501	0.359	0.75
BDUR	5772	-0.022	0.994	-0.698	-.149	0.497
GDP	6909	0.097	0.914	-0.648	0.124	1.236
Inflation	6909	0.335	1.050	0.028	0.709	1.495
Interest	6909	-0.238	0.433	-0.638	-0.499	0.269
MKT_p	6909	-0.618	1.324	-2.37	-0.129	0.531

Table 2
Pairwise correlations

Variables	CAR	RD	SD	SIZE	LEV	ROA	CapInt	MTB	MBE	Analysis	BIND	BDUR	GDP	Inflation	*Interest*	MKT_p
CAR	1.000															
RD	0.030**	1.000														
SD	-0.022*	-0.047***	1.000													
SIZE	-0.015	0.064***	-0.028**	1.000												
LEV	0.000	-0.073***	0.036***	0.029**	1.000											
ROA	-0.009	-0.023*	-0.167***	0.149***	0.011	1.000										
CapInt	-0.018	-0.099***	0.297***	0.006	0.065***	-0.092***	1.000									
MTB	0.002	0.001	-0.008	0.034***	-0.020*	0.052***	-0.001	1.000								
MBE	0.021*	-0.018	-0.029**	0.021*	-0.048***	0.076***	-0.077***	-0.004	1.000							
Analysis	0.019	0.097***	-0.077***	0.553***	0.020	0.047***	0.043***	0.027**	0.089***	1.000						
BIND	-0.051***	-0.041***	0.047***	0.119***	0.158***	-0.036***	-0.002	-0.052***	0.059***	0.060***	1.000					
BDUR	0.008	0.037***	-0.064***	-0.101***	-0.112***	0.114***	-0.042***	0.017	0.015	-0.106***	-0.348***	1.000				
GDP	0.025**	-0.011	0.113***	-0.066***	-0.025**	-0.046***	0.078***	0.027**	0.003	0.035***	-0.021	0.025*	1.000			
Inflation	0.019	-0.012	0.048***	-0.030**	-0.011	-0.013	0.052***	0.047***	-0.030**	0.011	0.010	0.033**	0.483***	1.000		
Interest	0.025**	-0.006	0.044***	-0.047***	-0.021*	-0.017	0.045***	0.031**	-0.006	0.032***	-0.010	0.050***	0.749***	0.490***	1.000	
MKT_p	-0.032***	-0.005	-0.059***	0.043***	0.020*	0.008	-0.051***	-0.048***	0.013	-0.032***	0.010	-0.037***	-0.710***	-0.537***	-0.825***	1.000

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3
Market effect of Space News on S&P500 Firms with High R&D Expenditure

	(1) CAR	(2) CAR	(3) CAR	(4) CAR
RD	0.062**	0.061**	0.057*	0.056*
	(0.035)	(0.041)	(0.096)	(0.097)
SD		-0.032	-0.000	-0.004
		(0.276)	(0.999)	(0.917)
SIZE		-0.026	-0.040	-0.033
		(0.236)	(0.160)	(0.247)
LEV		0.012	0.029	0.030
		(0.551)	(0.212)	(0.208)
ROA		-0.022	0.002	0.001
		(0.409)	(0.937)	(0.963)
CapInt		-0.023	-0.003	-0.006
		(0.377)	(0.907)	(0.844)
MTB		0.006	0.010	0.006
		(0.743)	(0.592)	(0.739)
MBE			0.040	0.042
			(0.464)	(0.438)
Analyst			0.005	0.005
			(0.200)	(0.271)
BIND			-0.091***	-0.091***
			(0.001)	(0.001)
BDUR			-0.011	-0.013
			(0.632)	(0.591)
GDP				0.000
				(0.993)
Inflation				-0.005
				(0.831)
Interest				-0.015
				(0.874)
MKT_p				-0.054*
				(0.074)
Constant	-0.032	-0.009	-0.180*	-0.203**
	(0.107)	(0.703)	(0.050)	(0.042)
Observations	6,909	6,343	5,226	5,226
R-squared	0.001	0.002	0.004	0.006

Robust p-val in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4
Market effect of launch failure on S&P500 Firms with High R&D Expenditure

	(1) CAR	(2) CAR	(3) CAR	(4) CAR
RD	0.131**	0.133**	0.113*	0.114*
	(0.035)	(0.033)	(0.093)	(0.091)
FAIL	0.063*	0.061	0.108**	0.106**
	(0.099)	(0.137)	(0.014)	(0.016)
*FAIL * RD*	-0.131**	-0.137**	-0.113*	-0.114*
	(0.034)	(0.028)	(0.094)	(0.089)
SD		-0.032	0.000	-0.004
		(0.278)	(0.998)	(0.917)
SIZE		-0.026	-0.040	-0.034
		(0.228)	(0.156)	(0.234)
LEV		0.012	0.031	0.030
		(0.537)	(0.188)	(0.198)
ROA		-0.025	0.000	0.000
		(0.357)	(0.994)	(0.988)
CapInt		-0.023	-0.004	-0.007
		(0.368)	(0.886)	(0.826)
MTB		0.006	0.010	0.006
		(0.731)	(0.596)	(0.746)
MBE			0.040	0.043
			(0.459)	(0.429)
Analyst			0.006	0.005
			(0.180)	(0.252)
BIND			-0.088***	-0.091***
			(0.000)	(0.001)
BDUR				-0.012
				(0.622)
GDP				-0.001
				(0.982)
Inflation				-0.008
				(0.752)
Interest				-0.017
				(0.862)
MKT_p				-0.056*
				(0.065)
Constant	-0.063	-0.039	-0.236**	-0.259**
	(0.100)	(0.346)	(0.014)	(0.013)
Observations	6,909	6,343	5,226	5,226

R-squared	0.002	0.003	0.006	0.007

Robust p-val in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5
Market effect of Space News on Aerospace Industry Firms with High R&D Expenditure

	(1) CAR	(2) CAR	(3) CAR	(4) CAR
RD	0.057*	0.054*	0.051	0.050
	(0.053)	(0.068)	(0.138)	(0.139)
AD	-0.409***	-0.406***	-0.304**	-0.305**
	(0.001)	(0.002)	(0.025)	(0.026)
*RD * AD*	1.265***	1.263***	1.734***	1.746***
	(0.005)	(0.005)	(0.007)	(0.006)
SD		-0.019	0.015	0.012
		(0.522)	(0.674)	(0.744)
Size		-0.019	-0.032	-0.025
		(0.376)	(0.254)	(0.374)
Leverage		0.013	0.033	0.033
		(0.530)	(0.168)	(0.164)
ROA		-0.024	-0.000	-0.001
		(0.364)	(0.995)	(0.970)
CapInt		-0.033	-0.015	-0.017
		(0.211)	(0.635)	(0.580)
MTB		0.006	0.010	0.006
		(0.739)	(0.598)	(0.744)
MBE			0.049	0.051
			(0.371)	(0.347)
Analyst			0.005	0.005
			(0.190)	(0.261)
BIND			-0.087***	-0.087***
			(0.001)	(0.001)
BDUR			-0.009	-0.010
			(0.709)	(0.662)
GDP				-0.001
				(0.989)
Inflation				-0.004
				(0.863)
Interest				-0.007
				(0.940)
MKT_p				-0.052*
				(0.084)
Constant	-0.022	-0.005	-0.186**	-0.206**
	(0.261)	(0.833)	(0.043)	(0.039)
Observations	6,909	6,343	5,226	5,226
R-squared	0.003	0.004	0.007	0.008

Robust p-val in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6
Market effect of launch failure R&D expenditure and aerospace industry

	(1) CAR	(2) CAR	(3) CAR	(4) CAR
*RD * AD*	2.565***	2.553***	3.148***	3.124***
	(0.003)	(0.003)	(0.002)	(0.002)
AD	-0.743***	-0.746***	-0.595**	-0.602***
	(0.001)	(0.001)	(0.010)	(0.010)
RD	0.120*	0.121*	0.101	0.102
	(0.053)	(0.053)	(0.134)	(0.131)
*AD * FAIL * RD*	-2.566***	-2.548***	-3.152***	-3.089***
	(0.003)	(0.004)	(0.002)	(0.003)
*AD * FAIL*	0.744***	0.752***	0.600***	0.614***
	(0.001)	(0.000)	(0.008)	(0.007)
*RD * FAIL*	-0.121*	-0.126**	-0.101	-0.102
	(0.052)	(0.043)	(0.136)	(0.130)
FAIL	0.045	0.040	0.088**	0.086*
	(0.245)	(0.328)	(0.048)	(0.055)
SD		-0.019	0.015	0.011
		(0.538)	(0.685)	(0.756)
SIZE		-0.019	-0.032	-0.026
		(0.384)	(0.250)	(0.357)
LEV		0.014	0.034	0.034
		(0.489)	(0.148)	(0.154)
ROA		-0.027	-0.002	-0.002
		(0.309)	(0.942)	(0.940)
CapInt		-0.034	-0.015	-0.017
		(0.203)	(0.625)	(0.573)
MTB		0.006	0.009	0.006
		(0.715)	(0.600)	(0.746)
MBE			0.048	0.051
			(0.372)	(0.347)
Analyst			0.006	0.005
			(0.182)	(0.253)
BIND			-0.085***	-0.087***
			(0.001)	(0.001)
BDUR				-0.009
				(0.696)
GDP				-0.003
				(0.945)
Inflation				-0.006
				(0.793)
Interest				-0.011

				(0.912)
MKT_p				-0.055*
				(0.071)
Constant	-0.045	-0.026	-0.230**	-0.251**
	(0.248)	(0.543)	(0.018)	(0.017)
Observations	6,909	6,343	5,226	5,226
R-squared	0.006	0.008	0.011	0.012

Robust p-val in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7
Market effect of R&D expenditure and analysis following

	(1)	(2)	(3)	(4)
	CAR	CAR	CAR	CAR
RD	0.109***	0.104***	0.113**	0.112**
	(0.007)	(0.009)	(0.016)	(0.016)
$Analyst_h$	0.018	0.027	0.033	0.026
	(0.644)	(0.597)	(0.547)	(0.633)
*RD * $Analyst_h$*	-0.113*	-0.106*	-0.159**	-0.158**
	(0.057)	(0.076)	(0.011)	(0.011)
SD		-0.031	-0.003	-0.007
		(0.302)	(0.923)	(0.844)
SIZE		-0.026	-0.019	-0.013
		(0.315)	(0.490)	(0.629)
LEV		0.012	0.029	0.029
		(0.539)	(0.218)	(0.215)
ROA		-0.023	-0.001	-0.001
		(0.389)	(0.980)	(0.960)
CapInt		-0.023	-0.002	-0.004
		(0.362)	(0.959)	(0.881)
MTB		0.007	0.010	0.006
		(0.687)	(0.589)	(0.738)
MBE			0.050	0.052
			(0.357)	(0.341)
BIND			-0.093***	-0.093***
			(0.001)	(0.001)
BDUR			-0.014	-0.015
			(0.553)	(0.520)
GDP				0.001
				(0.986)
Inflation				-0.005
				(0.826)
Interest				-0.009
				(0.926)
MKT_p				-0.053*
				(0.076)
Constant	-0.037	-0.020	-0.100*	-0.132*
	(0.139)	(0.580)	(0.076)	(0.056)
Observations	6,909	6,343	5,226	5,226
R-squared	0.002	0.003	0.005	0.007

Robust p-val in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8
Market effect of launch failure R&D expenditure and analysis following

	(1)	(2)	(3)	(4)
VARIABLES	CAR	CAR	CAR	CAR
RD	0.227***	0.223***	0.215**	0.215**
	(0.006)	(0.007)	(0.014)	(0.014)
Analyst$_h$	0.042	0.046	0.056	0.049
	(0.589)	(0.599)	(0.556)	(0.602)
*RD * Analyst$_h$*	-0.235*	-0.222*	-0.325**	-0.325**
	(0.059)	(0.075)	(0.010)	(0.010)
FAIL	0.079	0.071	0.124**	0.122**
	(0.111)	(0.185)	(0.031)	(0.033)
*FAIL * RD*	-0.227***	-0.228***	-0.212**	-0.214**
	(0.006)	(0.006)	(0.016)	(0.014)
*FAIL * Analyst$_h$*	-0.044	-0.033	-0.045	-0.045
	(0.570)	(0.691)	(0.616)	(0.614)
*FAIL * Analyst$_h$ * RD*	0.234*	0.222*	0.323**	0.325**
	(0.060)	(0.074)	(0.011)	(0.010)
SD		-0.030	-0.004	-0.007
		(0.315)	(0.913)	(0.839)
SIZE		-0.027	-0.018	-0.013
		(0.301)	(0.500)	(0.635)
LEV		0.013	0.030	0.030
		(0.517)	(0.209)	(0.206)
ROA		-0.027	-0.001	-0.002
		(0.318)	(0.961)	(0.939)
CapInt		-0.024	-0.002	-0.005
		(0.348)	(0.938)	(0.863)
MTB			0.053	0.054
			(0.328)	(0.314)
MBE		0.007	0.009	0.006
		(0.673)	(0.607)	(0.757)
BIND			-0.093***	-0.093***
			(0.001)	(0.001)
BDUR			-0.014	-0.015
			(0.565)	(0.533)
GDP				-0.001
				(0.986)
Inflation				-0.008
				(0.754)
Interest				-0.011
				(0.913)
MKT$_p$				-0.055*

				(0.067)
Constant	-0.077	-0.057	-0.164**	-0.196**
	(0.117)	(0.323)	(0.027)	(0.020)
Observations	6,909	6,343	5,226	5,226
R-squared	0.004	0.005	0.008	0.010

Robust p-val in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The Value of Corporate Natural Hazard Preparedness

(preliminary version)

Student Name: Ziye Huang
Supervisors: Ning Gao, Alice Liang Xu

1. Background and motivation

The increasing frequency and intensity of natural hazards have had significant consequences for businesses and communities worldwide. In recent years, natural hazards such as wildfires, hurricanes, and floods have caused severe disruptions, leading to financial losses, operational breakdowns, and reputational damage to companies. As an example, since January 7, 2025, a series of 31 wildfires have devastated Los Angeles, resulting in the loss of 27 lives, the evacuation of over 200,000 individuals, and the destruction of more than 12,401 structures. Such events underscore the critical need for companies to implement robust preparedness strategies to maintain operational stability and mitigate potential losses.

Corporate natural hazard preparedness is not only crucial during crises but also affects firm performance significantly. Firms that proactively manage hazard risks through effective preparedness tend to experience reduced operational disruptions and sustain profitability. Additionally, preparedness can influence firms' access to and cost of financing, particularly by affecting their perceived risk among lenders and investors. Despite its recognized importance, accurately evaluating corporate hazard preparedness from publicly available disclosures remains challenging.

In recent years, large language models (LLMs) have significantly advanced the potential of finance research, particularly in extracting meaningful insights from textual disclosures such as sustainability or corporate responsibility reports. LLMs are machine-learning models trained to process and generate human language. Traditional methods, like the bag-of-words approach, struggle to capture the meaning of text because they simplify textual data into word frequencies, ignoring the sequence of words and the surrounding context. Consequently, these methods fail to accurately capture the detailed meanings, relationships, and subtle contexts present in natural language. Advanced transformer-based LLMs, like ChatGPT, have overcome many limitations by better capturing context and relationships within textual data without extensive additional training. These models thus offer a promising tool for accurately evaluating and quantifying companies' hazard preparedness from their public sustainability or corporate responsibility reports.

The objective of this study is to explore how generative AI, particularly LLMs, can be leveraged to assess corporate readiness for natural hazards by analyzing corporate disclosures. This research is significant because it provides an innovative approach to evaluating natural hazard preparedness, which is crucial for sustaining business operations and protecting stakeholder interests. We aim to develop an AI-powered framework that uses large language models (LLMs) to assess the readiness of companies, identify key factors and impacts, and suggest ways to improve resilience. Moreover, this study aims to address the critical issue of assessing and enhancing corporate hazard preparedness, an area that remains inadequately addressed despite its acknowledged importance for businesses, the economy, and society, as highlighted in recent literature on climate risks and environmental challenges.

2. Hypothesis development

A firm's market value can be expressed as the present value of its expected future cash flows, discounted by its cost of capital (Modigliani and Miller, 1958). From this perspective, both the stability of cash flows and the level of risk faced by the firm play a central role in shaping valuation.

Natural hazards create significant operational and financial risks, disrupting business continuity, damaging assets, and generating costly recovery processes. Firms that lack adequate preparedness are more likely to suffer substantial cash flow losses and face longer, more expensive recovery periods. These risks increase the uncertainty of future earnings and raise the risk premium demanded by investors, ultimately resulting in lower market valuations relative to the book value of assets.

In contrast, firms that invest in natural hazard preparedness through explicit risk management strategies, crisis response plans, or broader resilience frameworks are better positioned to mitigate disruptions. Prepared firms are expected to maintain higher and more stable cash flows, reduce their exposure to costly losses, and recover more effectively when natural hazards occur. By lowering operational risk and the associated risk premium, preparedness translates into a higher present value of expected cash flows.

Accordingly, strong natural hazard preparedness should enhance firm value by both improving the stability of cash flows and reducing perceived risk. Based on this reasoning, we propose the following hypothesis:

H1. Corporate natural hazard preparedness is positively associated with firm value.

3. Variable construction

The central variable in this study is a firm-level Preparedness Index for natural hazards, derived from textual disclosures in sustainability and corporate responsibility reports. The construction of this variable follows a multi-stage pipeline that integrates dictionary-based context retrieval, context-preserving filtering, and large language model (LLM) evaluation.

The first stage involves employing a dictionary-based context retrieval procedure to identify hazard-related text within firm disclosures. To accomplish this, we construct a comprehensive dictionary that includes both general preparedness terminology (e.g., "disaster," "emergency," "crisis") and hazard-specific terms corresponding to the 19 FEMA-recognized natural hazard types (i.e., the 18 hazards included in the National Risk Index plus subsidence), along with their synonyms. This combined dictionary is applied to each report to extract sentences containing relevant terms.

In the second stage, we implement a context-preserving filtering approach to ensure that only hazard-relevant text is passed to the LLM. Specifically, the dictionary-based extraction from Stage 1 may still capture irrelevant sentences, since certain terms (e.g., "crisis") could refer to financial rather than natural hazard contexts. To address this, we feed the candidate sentences into GPT-5, which labels whether each sentence pertains to natural hazards. Sentences identified as relevant are retained, and to preserve contextual meaning we further expand the retained segments by including a five-sentence window before and after each match. This design improves both efficiency and accuracy: it prevents the model from processing entire corporate reports—which would be computationally costly and prone to hallucinations—while ensuring that evaluations remain focused on filtered, contextually grounded passages.

The final stage involves the evaluation of the filtered text using GPT-5 with prompt engineering specifically designed for preparedness assessment. At this stage, the model assigns a score to each firm based on the extent to which preparedness measures related to natural hazards are disclosed. We construct a three-level Preparedness Index: a score of 1 is assigned when the company explicitly mentions measures to address incidents caused by natural hazards or natural disasters; a score of 0.5 is assigned when the company refers only indirectly to such measures through broader frameworks such as “crisis management,” “emergency preparedness,” or “disaster recovery plans” that apply to all-hazards contexts; and a score of 0 is assigned when no preparedness measures are mentioned. The resulting index transforms unstructured disclosures into a structured, quantitative measure of corporate readiness for natural hazards, enabling systematic comparison of preparedness across firms and industries.

The final dataset comprises 9,991 firm-year observations. Table 1 provides a summary definition of the Preparedness Index.

Table 1: Definition of the preparedness index

Score	Definition	Example Disclosure
1 (Direct)	The company explicitly mentions measures to address incidents caused by natural hazards.	"We maintain a business continuity plan and disaster recovery plan to ensure operations during floods or hurricanes."
0.5 (Indirect)	The company does not explicitly mention measures to address incidents caused by natural hazards, but refers to broader ones such as "crisis management," "emergency preparedness," or "disaster recovery plans," in a context that applies to all-hazards crises, disasters, or emergencies.	"Our Business Continuity Plan (BCP) includes emergency response plans with an all-hazards approach to cover a wide range of potential threats and risks."
0 (None)	No disclosure of preparedness measures.	“The company does not reference any measures related to natural hazards.”

Figure 1 presents the distribution of firm-year observations in the sample from 2002 to 2023. The number of observations rises steadily over time, reflecting the increasing availability of sustainability and corporate responsibility disclosures in recent years. To ensure data quality, we consider two sample periods in the main regression: 2015–2023, which ensures at least 400 observations per year, and 2013–2023, which ensures at least 200 observations per year.

Figure 1: Distribution of firm-year observations (2002–2023)

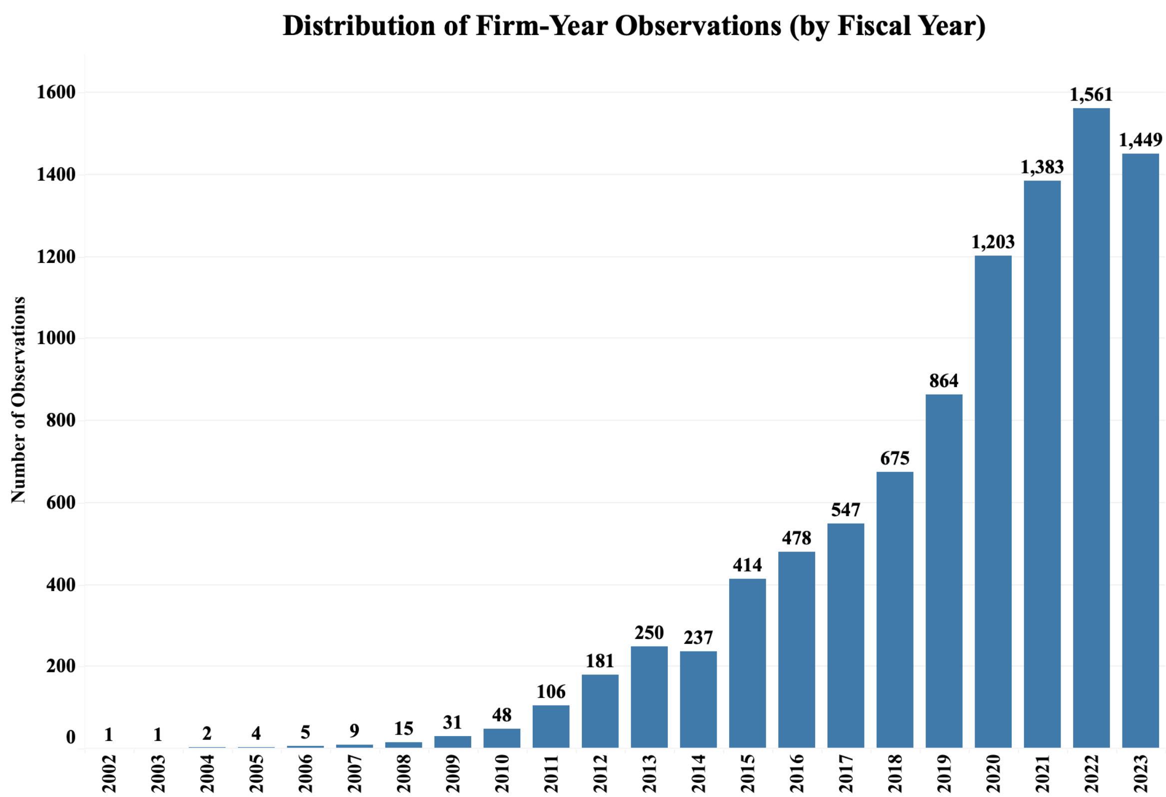


Figure 2 shows the proportion of firm-year observations classified as direct, indirect, or none between 2015 and 2023. From this figure, we can see that direct mentions of measures to incidents caused by natural hazards increase over time, while the proportion of firms with no relevant measures declines, indicating a gradual improvement in preparedness measures for natural hazards.

Figure 2: Proportion of direct, indirect, and none labels per fiscal year (2015–2023)

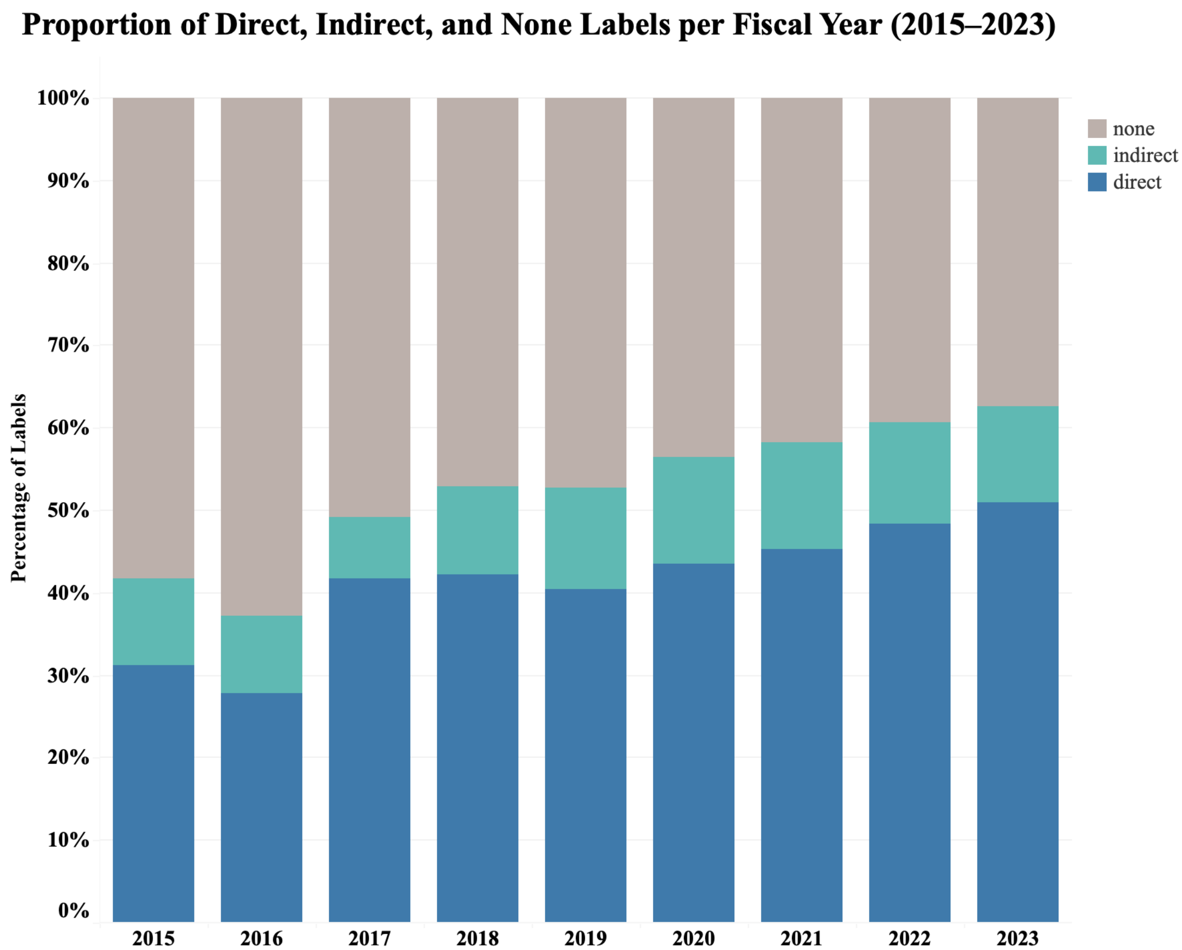
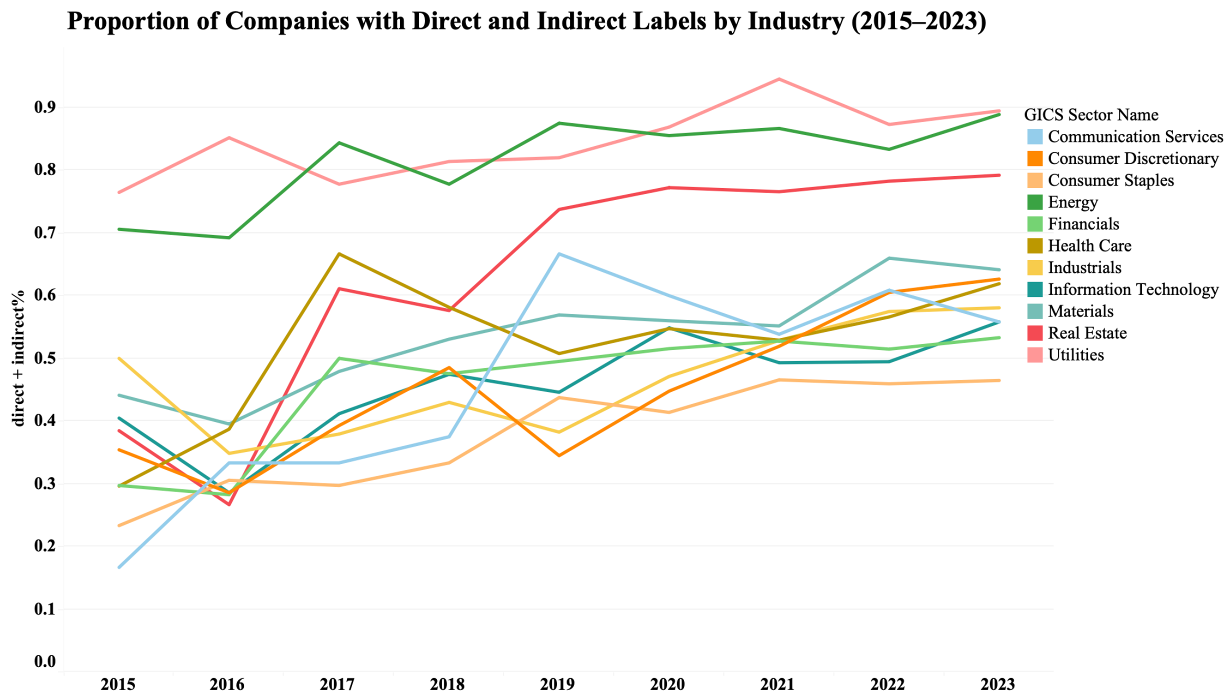


Figure 3 shows the proportion of companies with direct or indirect preparedness labels across GICS industries between 2015 and 2023. The results indicate substantial cross-industry variation, with Utilities, Real Estate, and Energy consistently showing a larger share of firms mentioning natural hazard preparedness measures, while other industries display more gradual increases over time.

Figure 3: Proportion of companies with direct and indirect labels by industry (2015–2023)



Figures 4 and 5 compare the geographic distribution of companies with direct or indirect labels across U.S. states in 2015 and 2023. The results show a clear increase in the number of companies reporting natural hazard preparedness measures over time, with growth observed across nearly all states. The concentration of such companies is particularly evident in coastal states such as California, Texas, Florida, and New York, where exposure to natural hazards is relatively high.

Figure 4: Number of companies with direct or indirect labels for natural hazard emergency management by state (2015)

Number of Companies with Direct or Indirect Label for Natural Hazard Emergency Management by State (2015)

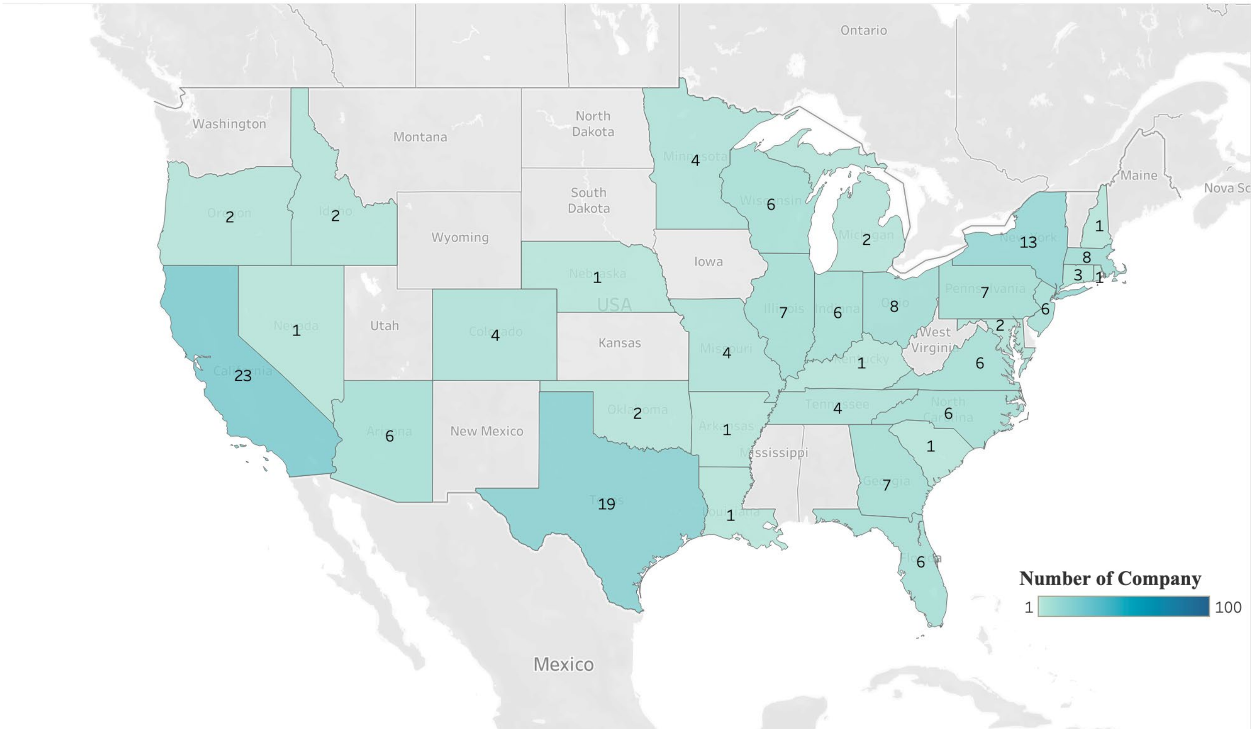
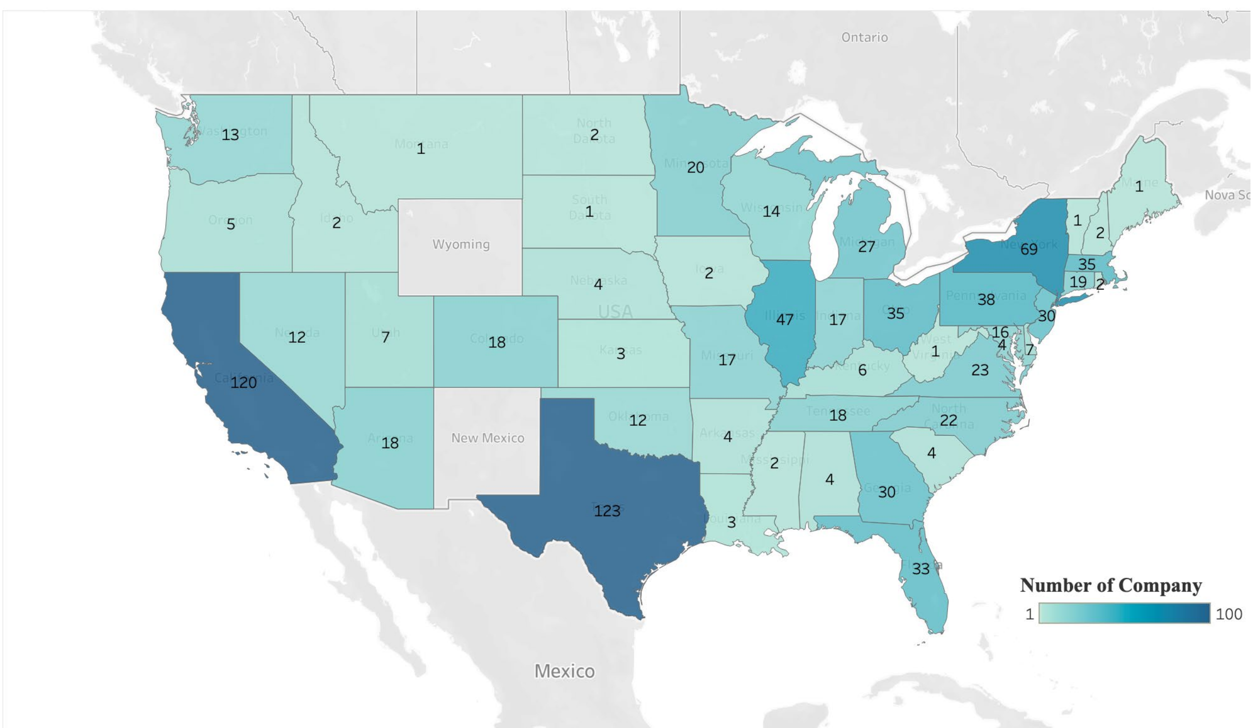


Figure 5: Number of companies with direct or indirect labels for natural hazard emergency management by state (2023)

Number of Companies with Direct or Indirect Label for Natural Hazard Emergency Management by State (2023)



4. Empirical results

4.1 Descriptive statistics

Table 2 presents the descriptive statistics of the main variables used in the analysis. The dependent variable in the main regressions, Tobin's Q, has a mean of 1.821 and a median of 1.021, indicating that the average firm in the sample is valued above the book value of its assets, though the distribution is right-skewed with a relatively high standard deviation (3.189). The Preparedness Index variable (Label) has a mean of 0.479, suggesting that around half of firms have mentioned measures to address incidents caused by natural hazards in their public disclosures. Control variables exhibit expected distributions: firms report an average return on assets (ROA) of 3.9 percent, median leverage of 0.324, and average firm age of about 9.5 years (log value ≈ 2.25). The average volatility of ROA is 0.050, indicating that firms' profitability typically fluctuates by about five percentage points across years.

Table 2: Descriptive statistics for main variables (N = 5,748)

Variables	Mean	SD	P25	Median	P75
Tobin's Q (t+1)	1.821	3.189	0.585	1.021	1.932
Label	0.479	0.468	0.000	0.500	1.000
ROA	0.039	0.097	0.011	0.042	0.084
Log(Firm Size)	8.862	1.551	7.739	8.791	9.909
Leverage	0.336	0.191	0.205	0.324	0.450
Growth	0.109	0.258	-0.013	0.067	0.180
Log(Firm Age)	2.251	0.288	2.079	2.303	2.485
Tangibility	0.290	0.255	0.085	0.192	0.465
σ(ROA)	0.052	0.083	0.015	0.029	0.057

Note: This table reports the summary statistics for variables used in the main tests. Detailed variable definitions are provided in Appendix 1.

4.2 Regression results on market valuation

Table 3 reports the regression of the preparedness index (Label) on firms' forward market valuation, measured by Tobin's Q (t+1). In Column (1), the analysis is based on the sample period 2015–2023, which ensures data quality with at least 400 observations per year. The coefficient on Label is positive (0.194) and statistically significant at the 5% level, supporting the hypothesis that disaster preparedness enhances firm value. Column (2) expands the sample period to 2013–2023, while maintaining a minimum of 200 observations per year. The coefficient on Label remains positive (0.183) and statistically significant at the 5% level, confirming that the results are robust to alternative time windows.

Both specifications include state and year fixed effects, which control for regional and time-specific factors. We do not include firm fixed effects because the preparedness index changes little within firms over time. In practice, a firm that is well prepared in one year is usually also well prepared in the next. Since the index is highly persistent, adding firm fixed effects would remove most of its variation and make it difficult to estimate its impact on valuation.

Overall, the evidence consistently suggests that corporate preparedness for natural hazards contributes to improved market valuation.

Table 3: Impact of Hazard Preparedness on Firm Market Valuation (Tobin's Q)

Variables	Dependent Variable: Tobin's Q	
	(1)	(2)
Label	0.194**	0.183**
	(2.05)	(2.06)
ROA	8.082***	8.286***
	(6.38)	(6.47)
Log(Firm Size)	-0.134*	-0.144*
	(-1.75)	(-1.82)
Leverage	0.730	0.699
	(1.23)	(1.24)
Growth	0.802*	0.791*
	(1.86)	(1.86)
Log(Firm Age)	-0.620*	-0.622*
	(-1.76)	(-1.86)
Tangibility	-2.514***	-2.338***
	(-6.50)	(-6.78)

σ(ROA)	1.315	1.447
	(0.81)	(0.92)
Fixed effects:		
State	YES	YES
Year	YES	YES
No. of observations	5748	6166
Adjusted	0.142	0.148

Note: This table presents the regression results of the relationship between the preparedness index (Label) and firm market valuation, measured by Tobin's Q. Column (1) covers the period 2015–2023, requiring at least 400 firm-year observations per year. Column (2) expands the sample to 2013–2023, requiring at least 200 firm-year observations per year. Robust standard errors are clustered at the state level. *p < 0.10, **p < 0.05, and ***p < 0.01. Detailed variable definitions are provided in Appendix 1.

4.3 Preparedness index and firm valuation: robustness check with group sorting

To check the robustness of the results, firms are sorted into High, Medium, and Low preparedness groups. Low-preparedness firms serve as the benchmark. The results in Table 4 show that High-preparedness firms have significantly higher market valuation, with a high–low spread of 0.183(t = 2.08, significant at the 5% level). Medium-preparedness firms also display a positive, though statistically insignificant, spread of 0.088 (t = 0.65). This indicates that the valuation premium is concentrated among the most prepared firms. All regressions include state and year fixed effects, and standard errors are clustered at the state level. The High–Low spread in Tobin's Q remains positive and significant, supporting the main finding that preparedness is valued by the market.

Table 4: Preparedness index and firm valuation (group sorting approach)

Variables	Dependent Variable: Tobin's Q
High preparedness	0.183**
	(2.08)
Medium preparedness	0.088
	(0.65)
ROA	8.286***
	(6.47)
Log(Firm Size)	-0.144*
	(-1.82)
Leverage	0.699
	(1.25)
Growth	0.791*
	(1.86)
Log(Firm Age)	-0.622*

	(-1.87)
Tangibility	-2.338***
	(-6.75)
σ(ROA)	1.449
	(0.92)
Fixed effects:	
State	YES
Year	YES
No. of observations	6166
Adjusted	0.147

Note: This table presents the regression results of the relationship between the preparedness index and firm market valuation, measured by Tobin's Q. Firms are classified into High, Medium, and Low preparedness groups, with Low preparedness serving as the benchmark. Robust standard errors are clustered at the state level. *p < 0.10, **p < 0.05, and ***p < 0.01. Detailed variable definitions are provided in Appendix 1.

4.4 Preparedness index and operating cash flow return on assets

Table 5 examines the effect of hazard preparedness on operating cash flow return on assets. Consistent with our hypothesis, the coefficient on the preparedness index is positive and statistically significant at the 5% level (0.004, t = 2.04), suggesting that firms with stronger preparedness generate higher operating profitability. Economically, a one-unit increase in preparedness corresponds to a 40 basis point increase in operating cash flow returns. Among the controls, industry concentration (HHI) is positively associated with profitability, while firm size, leverage, growth opportunities, and cash flow volatility are not significant. These results support the view that natural hazard preparedness enhances firms' ability to maintain higher operating cash flows by mitigating hazard-related disruptions, thereby strengthening financial resilience and contributing to greater firm value.

Table 5: Preparedness index and operating cash flow return on assets

Variables	Dependent Variable: Operating Cash Flow Return on Assets
Label	0.004**
	(2.04)
Log(Firm Size)	-0.001
	(-0.60)
Leverage	0.018
	(1.13)
Growth	0.015
	(1.20)

σ(ROA)	-0.070
	(-0.86)
HHI	0.000**
	(2.38)
Fixed effects:	
State	YES
Year	YES
No. of observations	8489
Adjusted	0.043

Note: This table presents the regression results of the relationship between hazard preparedness and the operating cash flow return on assets. The key independent variable is the preparedness index (Label). Regression includes state and year fixed effects. Robust standard errors are clustered at the state level. *p < 0.10, **p < 0.05, and ***p < 0.01. Detailed variable definitions are provided in Appendix 1.

4.5 Preparedness index and cash flow volatility

In line with our hypothesis (H1), we argue that preparedness enhances firm value by reducing the volatility of operating cash flows, thereby stabilizing earnings and lowering perceived risk. To test this mechanism, Table 6 presents regressions of the volatility of operating cash flow return on assets on the preparedness index. The coefficient on preparedness is negative and statistically significant at the 5% level (–0.002, t = –2.31), indicating that better-prepared firms experience more stable operating cash flows. In economic terms, a one-unit increase in preparedness is associated with a 20 basis point reduction in cash flow volatility. This result suggests that preparedness reduces firms’ exposure to hazard-induced disruptions, shortens recovery periods, and mitigates costly operational shocks. By improving the stability of cash flows, preparedness not only lowers firms’ operational risk but also reduces the risk premium demanded by investors. Taken together, this evidence supports the hypothesized channel, indicating that natural hazard preparedness contributes to higher firm valuation by stabilizing operating cash flows.

Table 6: Preparedness index and volatility of operating cash flow return on assets

Variables	Dependent Variable: Volatility of Operating Cash Flow Return on Assets
Label	-0.002**
	(-2.31)
ROA	0.008
	(0.69)
Log(Firm Size)	-0.006***
	(-17.80)

Leverage	0.002
	(0.63)
Growth	0.023***
	(6.24)
Log(Firm Age)	-0.012***
	(-4.67)
Fixed effects:	
State	YES
Year	YES
No. of observations	8595
Adjusted	0.236

Note: This table presents the regression results of the relationship between hazard preparedness and the volatility of operating cash flow return on assets. The key independent variable is the preparedness index (Label). Regression includes state and year fixed effects. Robust standard errors are clustered at the state level. *p < 0.10, **p < 0.05, and ***p < 0.01. Detailed variable definitions are provided in Appendix 1.

4.6 Preparedness index and operating cash flow growth

We further examine whether preparedness contributes to firm value by supporting stronger operating performance. Specifically, we test whether the preparedness index is associated with higher growth in operating cash flows, as preparedness reduces business disruptions and limits losses from natural hazards. Table 7 reports the regression results, where the dependent variable is the operating cash flow growth rate. The coefficient on preparedness is positive and statistically significant at the 5% level (0.070, t = 2.55), indicating that better-prepared firms experience faster growth in operating cash flows. Economically, a one-unit increase in preparedness is associated with a 7.0 percentage point increase in the annual growth rate of operating cash flows. This evidence suggests that preparedness not only stabilizes firms against downside shocks but also helps preserve growth opportunities by mitigating disruption losses and maintaining operational continuity. Overall, these results provide complementary support for the expected cash flow channel: better-prepared firms generate higher expected cash flows, which the market capitalizes into higher firm value.

Table 7: Preparedness index and operating cash flow growth rate

Variables	Dependent Variable: Operating Cash Flow Growth Rate
Label	0.070**
	(2.55)
Log(Firm Size)	-0.017

	(-1.13)
Leverage	0.059
	(0.68)
Log(Firm Age)	-2.149***
	(-4.44)
σ(ROA)	0.149**
	(2.01)
Tangibility	0.066
	(1.15)
Fixed effects:	
State	YES
Year	YES
No. of observations	7982
Adjusted	0.006

Note: This table reports regressions of the relationship between hazard preparedness and the growth rate of operating cash flows. The key independent variable is the preparedness index (Label). All regressions include state and year fixed effects. Robust standard errors are clustered at the state level. *p < 0.10, **p < 0.05, ***p < 0.01. Detailed variable definitions are provided in Appendix 1.

5. Future Empirical Tests

In order to strengthen the robustness of our findings, several further tests will be conducted. First, we will examine additional determinants of firm performance, including market performance during crisis periods and the cost of debt. Second, we will refine the natural hazard preparedness index by controlling for firms' actual exposure to natural hazards and by adjusting for systematic differences across industries and geographic locations. Finally, we plan to implement a validation test through the analysis of company-level losses incurred during and after hazard events. This will allow us to directly evaluate whether preparedness levels are associated with reduced realized losses, thereby reinforcing the practical relevance of the preparedness index as an indicator of resilience.

Appendix 1: Variable description

Variable	Description
Tobin's Q	Ratio of market value of assets to book value of assets.
Label	Preparedness index. Equals 1 for high preparedness, 0.5 for medium preparedness, 0 for low preparedness.
ROA	Return on assets, measured as operating cash flow divided by total assets.
Log(Firm Size)	Natural logarithm of total assets.
Leverage	Ratio of total debt to total assets.
Growth	Annual sales growth rate.
Log(Firm Age)	Natural logarithm of firm age in years.
Tangibility	Ratio of tangible assets to total assets.
σ(ROA)	Standard deviation of return on assets (proxy for profitability volatility).
HHI	Industry concentration measured by firms' asset shares.

Reference:

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American economic review*, *48*(3), 261-297.

REGULATION, MARKET FRAGMENTATION, AND STARTUP FINANCING: EVIDENCE FROM EUROPEAN VENTURE CAPITAL

Fábio Sousa (fabio.sousa@manchester.ac.uk)
Alliance Manchester Business School, University of Manchester

Amedeo De Cesari (amedeo.decesari@manchester.ac.uk)
Alliance Manchester Business School, University of Manchester

Konstantinos Stathopoulos (k.stathopoulos@manchester.ac.uk)
Alliance Manchester Business School, University of Manchester
Copenhagen Business School

This version: March 20, 2026

ABSTRACT

Academics and policymakers increasingly worry about venture capital (VC) concentration, arguing that a small number of investors may control a disproportionate share of funding and act as gatekeepers in entrepreneurial finance. Yet whether more competitive VC markets improve entrepreneurial financing is theoretically ambiguous. Venture capital is characterized by severe information asymmetries. While greater concentration may mitigate these frictions by enhancing screening and monitoring through scale and experience, it may also reduce competitive pressure and limit the supply of capital. This paper studies how VC market structure affects startup financing. We exploit the implementation of the Alternative Investment Fund Managers Directive (AIFMD) as a quasi-natural experiment that altered the European VC landscape. We show that the AIFMD reduced market concentration. However, this increase in competition is associated with weaker financing outcomes: startups raise less capital, attract fewer investors, and receive smaller investments per investor relative to the counterfactual. These findings highlight a trade-off between competition and efficiency in VC markets and suggest that policies aimed at reducing concentration may have unintended consequences for entrepreneurial finance.

JEL (G24, G28, L11)

Keywords: Venture Capital; Market Concentration; Entrepreneurial Finance; Financial Regulation; AIFMD; Startup Financing

1. INTRODUCTION

Venture capital is a cornerstone of modern innovation ecosystems. Venture capital investors financed some of the most transformative companies of the past decades and play a critical role in the commercialization of new technologies (Lerner and Nanda 2020). Prior research shows that venture capital funding stimulates innovation (e.g. Lindsey (2008); Puri and Zarutskie (2012); Kerr, Lerner and Schoar (2014); González-Uribe (2020)), accelerates firm growth (e.g. Inderst and Mueller (2009); Chemmanur, Krishnan and Nandy, (2014)), and improves the performance of firms going public (e.g. Megginson and Weiss (1991); Lerner (1994); Brav and Gompers (1997); Sørensen (2007); Nahata (2008)). As a result, the availability of venture capital has become a key determinant of entrepreneurial dynamism and technological progress.

Despite its central role in financing innovation, relatively little is known about the industrial organization of the venture capital industry. In particular, the extent to which the market structure of venture capital markets influences startup outcomes remains largely unexplored. This question is especially important given growing concerns about the increasing concentration of venture capital investment (Lerner and Nanda 2020; The Purposeful Company 2025). Recent evidence shows that a small number of large investors control a substantial share of the capital deployed to startups. For example, Lerner and Nanda (2020) document that roughly five percent of venture capital firms accounted for nearly half of total capital raised in the United States during the 2010s. Such concentration implies that a relatively small set of investors may effectively act as gatekeepers in entrepreneurial finance, influencing which startups receive funding and which technologies are ultimately developed (Lerner and Nanda 2020; The Purposeful Company 2025).

Whether concentration in venture capital markets is beneficial or harmful for entrepreneurial activity is theoretically ambiguous. On the one hand, traditional industrial organization theory predicts that greater competition among investors may improve access to finance by expanding the supply of capital and lowering financing costs. On the other hand, venture capital investments involve substantial information frictions and require intensive screening, monitoring, and governance (e.g. Gompers (1995); Kaplan et al (2001); Winton and Yerramilli (2008)). In such settings, (see Petersen and Rajan (1995) for banking) increased market power might lead to larger and more established investors with superior expertise, networks, and deal flow, potentially improving the allocation of capital (e.g. Hochberg et al (2007)). These opposing mechanisms suggest that the relationship between venture capital market concentration and startup financing cannot be determined *a priori*.

While a large literature studies the effects of market structure in banking and other forms of financial intermediation (e.g. Massa (2003); Wahal and Wang (2011); Khorana and Servaes (2012); Petersen and Rajan (1995); Cetorelli and Gambera (2001); Cetorelli and Strahan (2006), Kerr and Nanda (2009), and Rice and Strahan (2010)), comparable evidence for venture capital markets remains scarce. Existing research has primarily focused on how the supply of venture capital determines the geographic distribution of startup activity (e.g. Chen and Ewens (2025)), rather than the consequences of concentration among venture capital investors themselves. Consequently, we still know relatively little about how changes in the competitive structure of venture capital markets influence the availability of entrepreneurial finance.

In this paper, we address this question by studying how venture capital market concentration affects startup financing outcomes. Identifying the causal impact of market concentration is challenging because venture capital market structure is typically endogenous to local economic conditions and entrepreneurial activity. To overcome this challenge, we exploit the implementation of the Alternative Investment Fund Managers Directive (AIFMD) as a natural experiment that altered the regulatory environment of venture capital in Europe (European Commission et al. 2025).

Introduced after the Global Financial Crisis, the AIFMD established a comprehensive regulatory framework for alternative investment fund managers in the European Union. The Directive imposed a range of requirements - including capital standards, disclosure obligations, risk management procedures, and ongoing regulatory reporting - that significantly increased the fixed costs of operating a fund (Clifford Chance 2012; Qadir 2016; European Commission et al. 2025). Evidence from the European Commission suggests that ongoing compliance costs can represent up to 20 percent of a fund's operational expenses, making them economically significant, particularly for smaller managers (European Commission et al. 2025).

A key feature of the AIFMD is its threshold-based design. The full regulatory regime applies only to managers exceeding €100 million in assets under management (or €500 million for unleveraged funds with long lock-up periods), while smaller managers face substantially lighter requirements (Clifford Chance 2012). This creates a strong incentive for fund managers to remain below the compliance thresholds, as crossing them triggers a discrete increase in regulatory burden and costs (European Commission et al. 2025).

These threshold effects are particularly relevant in venture capital markets, where funds are typically small, operate with tight margins, and rely on back-loaded compensation structures (European Commission et al. 2025). For many managers, scaling up fund size may imply a disproportionate increase in compliance costs relative to expected returns (European Commission et al. 2025). As a result, the AIFMD may discourage fund growth and induce managers to remain small or to split operations across multiple vehicles in order to avoid triggering full regulatory compliance (European Commission et al. 2025).

In addition, while the Directive introduced an EU-wide passport to facilitate cross-border activity, access to this regime requires full compliance. Smaller managers that remain below the thresholds cannot fully benefit from these integration gains, while cross-country differences in implementation ("gold-plating") further increase regulatory complexity (European Commission et al. 2025).

Taken together, the AIFMD introduces a set of institutional frictions that can reshape the market structure of venture capital. While higher compliance costs could, in principle, favor larger incumbents, the threshold-based design and associated distortions instead create incentives for fragmentation. In this paper, we empirically test how this regulatory shock affected venture capital market concentration and examine the consequences for startup financing.

We use this regulatory change to study how shifts in venture capital market structure affect startup financing. Using investment-level data from LSEG Workspace covering venture capital investments across major economies between 2008 and 2018, we proceed in two steps. First, we examine whether the implementation of the AIFMD affected concentration levels in European venture capital markets. Second, we analyze how these changes in market structure influenced startup financing outcomes.

Our results provide evidence that the AIFMD significantly reduced venture capital market concentration in Europe. Using difference-in-differences specifications and alternative transformations of the Herfindahl–Hirschman Index (HHI), we find that concentration declined substantially following the implementation of the Directive. This decline is consistent with the interpretation that regulatory compliance costs and institutional frictions led to greater fragmentation among venture capital investors.

We then examine the implications of this change in market structure for startup financing. Our findings indicate that the reduction in concentration is associated with weaker financing outcomes for startups. Following the implementation of the AIFMD, startups in treated European markets raise significantly less capital, attract fewer investors per funding round, and receive smaller investments from each participating investor relative to comparable startups in the United States.

These results suggest that fragmentation in venture capital markets may reduce the availability of entrepreneurial finance. In particular, regulatory policies that increase compliance costs or discourage the scaling of venture capital funds may unintentionally constrain the supply of capital available to startups.

This paper contributes to several strands of literature. First, we contribute to the growing literature on the industrial organization of financial intermediation by examining how market structure affects financing outcomes. Second, we extend the venture capital literature by studying the vertical consequences of concentration among venture capital investors. Finally, we contribute to the literature on the real effects of financial regulation by showing how regulatory changes can reshape the structure of venture capital markets and influence entrepreneurial finance.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1. Literature Review

A large literature examines the industrial organization of financial intermediation.

A growing body of work studies concentration and competition in the asset management industry. Important contributions include Massa (2003), Wahal and Wang (2011), Khorana and Servaes (2012), Hoberg et al (2018), Feldman et al (2020), Kostovetsky and Warner (2020), Sun (2021), Loseto and Mainardi (2024), and Malenko et al. (2026). Related research analyzes how regulatory or policy shocks affect competition among financial intermediaries. For example, Giambona et al. (2025) and Bellia et al. (2025) study how policy interventions shape competition among insurance companies and designated market makers, respectively.

More closely related to our study is the literature examining how the market structure of financial intermediaries affects the outcomes of the firms they finance. This strand of research is particularly rich in studies investigating how bank concentration influences firm-level outcomes, including access to entrepreneurial finance.

One stream of this literature argues that higher market concentration can improve financing outcomes for small and young firms. Classic contributions include Petersen and Rajan (1995) and Cetorelli and Gambera (2001). Petersen and Rajan (1995) develop a theoretical framework based on relationship lending. In their model, banks with greater market power can sustain long-term lending relationships by subsidizing young or financially constrained firms early and recouping profits later. Because young firms typically lack credit histories and verifiable information, these informational frictions pose a major barrier to external financing (Petersen and Rajan 1995). Relationship lending allows banks to accumulate borrower-specific information over time, reducing information asymmetries that often hinder lending to small and opaque firms (Petersen and Rajan 1995). In competitive credit markets, however, lenders cannot easily capture future rents from successful borrowers (Petersen and Rajan 1995). As a result, they have weaker incentives to finance young or informationally opaque firms (Petersen and Rajan 1995). Using survey data from the United States, Petersen and Rajan (1995) provide empirical support for these predictions. Cetorelli and Gambera (2001) find that increased bank market power promotes industrial growth by expanding the amount of credit that is available to young firms.

A second stream of research reaches the opposite conclusion, arguing that greater banking competition improves access to entrepreneurial finance. Evidence from Black and Strahan (2002), Cetorelli (2004), Beck et al (2004), Cetorelli and Strahan (2006), Kerr and Nanda (2009), and Rice and Strahan (2010) suggests that reductions in banking concentration tend to improve funding conditions for startups, either by expanding credit supply or lowering borrowing costs. These conclusions appear to hold across several institutional settings. However, Beck et al (2004) show that the relationship between bank competition and credit access depends on the institutional environment. Legal frameworks, creditor protection, and financial regulation shape how banking competition translates into credit availability for firms (Beck et al. 2004). Kerr and Nanda (2009) further show that reductions in banking concentration stimulate entrepreneurship but also increase "churning" among new firms, as many young businesses fail early in their lifecycle. These findings suggest that improved access to finance expands the pool of potential entrepreneurs rather than necessarily increasing the success rate of new ventures (Kerr and Nanda 2009).

Core (2025) also fits within this second stream but offers a distinct perspective on how bank market power interacts with entrepreneurship policy. The author examines how banking market structure affects the effectiveness of policies designed to promote innovative entrepreneurship (Core 2025). Exploiting the introduction of the Start-Up Italy Act (SIA) of 2012 - which provided generous public guarantees on bank loans to innovative startups - Core (2025) shows that policy effectiveness depends critically on the competitiveness of local banking markets. While the reform substantially increased overall startup creation, the effects were concentrated in provinces with more competitive banking sectors (Core 2025). Core (2025)'s results indicate that bank market power weakens the transmission of policy incentives. In concentrated banking markets, banks pass through fewer credit subsidies, leading to lower volumes of guaranteed lending (Core 2025).

di Patti and Dell'Ariccia (2004) bridge these two views by showing that the relationship between bank concentration and entrepreneurial finance may be nonlinear. Using detailed local-level data from Italy across 22 industries and 103 provinces, the authors proxy credit availability using firm birth rates (di Patti and Dell'Ariccia 2004). They find that moderate increases in bank concentration can stimulate firm creation by strengthening incentives for relationship lending and information acquisition (di Patti and Dell'Ariccia 2004). However, beyond a certain threshold, further increases in concentration reduce firm creation, consistent with the conventional view that excessive market power restricts credit supply (di Patti and Dell'Ariccia 2004).

Our study contributes directly to the venture capital literature. Venture capital plays a central role in entrepreneurial finance (Lerner and Nanda 2020). Prior research shows that venture capital funding stimulates innovation (e.g. Lindsey (2008); Puri and Zarutskie (2012); Kerr, Lerner and Schoar (2014); González-Uribe (2020)), supports the growth of competitive startups (e.g. Inderst and Mueller (2009); Chemmanur, Krishnan and Nandy, (2014)), and improves IPO outcomes (e.g. Megginson and Weiss (1991); Lerner (1994); Brav and Gompers (1997); Sørensen (2007); Nahata (2008)).

Despite the central role of venture capital in entrepreneurial ecosystems, the literature examining the vertical effects of venture capital concentration remains limited.

Lerner and Nanda (2020) document a high degree of concentration in the venture capital industry, where a relatively small number of firms control a large share of the capital deployed to startups. Examining institutional venture capital investors in the United States, they show that roughly five percent of VC firms accounted for about half of the total capital raised between 2014 and 2018 (Lerner and Nanda 2020). This concentration implies that a small group of large investors plays a disproportionate role in allocating risk capital and shaping the direction of technological innovation (Lerner and Nanda 2020). The authors also show that venture capital investment has increasingly concentrated in a narrow set of sectors, particularly software and digital services (Lerner and Nanda 2020). This pattern reflects the economic incentives faced by venture capitalists, who tend to invest in sectors where uncertainty about market demand can be resolved quickly and successful exits occur within the typical venture fund lifecycle (Lerner and Nanda 2020). As a result, while venture capital can efficiently finance certain types of innovation, it may underfund technologies with longer development horizons - such as renewable energy and advanced materials - that may generate substantial social value (Lerner and Nanda 2020). Recent evidence from Bonelli (2025) reinforces this concern, showing that venture capitalists tend to fund startups that are technologically similar to those already in their portfolios, which may further narrow the range of innovations receiving financing.

Chen and Ewens (2025) provide related evidence by examining how venture capital supply shapes the geographic clustering of high-growth startups in the United States. Using the Volcker

Rule as an exogenous shock to venture capital fundraising, the authors show that regions more exposed to the withdrawal of bank limited partner capital experienced significantly smaller funds - declining by approximately 20–22 percent - and lower probabilities of follow-on fundraising (Chen and Ewens 2025). These effects were strongest in non-hub regions where venture capital markets were already relatively underdeveloped (Chen and Ewens 2025). As a result, startups in more exposed regions raised smaller funding rounds (about 7 percent smaller), received lower valuations (approximately 9% lower), and were more than 30 percent more likely to relocate to established venture capital hubs such as California (Chen and Ewens 2025). The authors also show that distant venture capitalists do not fully substitute for local investors, highlighting the importance of local venture capital supply and monitoring (Chen and Ewens 2025).

Concerns about concentration in venture capital markets also appear in policy discussions. The Purposeful Company - a coalition of leading academics and business leaders focused on the development of UK capital markets - highlights similar patterns of concentration within the venture capital ecosystem (The Purposeful Company 2025). The report notes that a small number of investors account for a disproportionate share of capital deployment and may act as gatekeepers within entrepreneurial finance (The Purposeful Company 2025). Because these investors possess substantial financial resources and frequently participate in later-stage financing rounds, they can effectively determine which startups receive backing and scale (The Purposeful Company 2025). In doing so, they shape the overall direction of venture capital investment, influencing which technologies, sectors, and business models receive funding (The Purposeful Company 2025).

Another concern relates to diversity within the venture capital ecosystem. The Purposeful Company (2025) highlights that venture capital firms are disproportionately dominated by graduates of elite institutions and remain underrepresented by women and minorities. Prior research documents strong homophily in venture capital networks and shows that underrepresented founders - particularly women - face greater barriers in accessing venture capital financing (e.g. Ewens and Townsend (2020); Calder-Wang and Gompers (2021)). A growing literature further shows that diversity among decision-makers and entrepreneurs improves economic outcomes (e.g. Huang and Kisgen (2013); Ewens and Townsend (2020); Calder-Wang and Gompers (2021); Francis et al (2021)). Consequently, higher levels of venture capital concentration could reduce the diversity of funded founders, potentially leading to suboptimal social and economic outcomes.

Our paper focuses explicitly on the market structure of the venture capital industry, thus making a clear contribution for both the literature and policy making. Chen and Ewens (2025) do not directly examine what happens when a small number of venture capital investors hold substantial market power within the market. In contrast, we study how regulatory shocks affect concentration among venture capital investors and examine the implications for startup financing. By analyzing how regulation alters the competitive landscape of venture capital markets, this study contributes to understanding whether concentration among venture capital intermediaries influences which startups, technologies and entrepreneurs receive funding and how risk capital is allocated within the entrepreneurial ecosystem.

2.2. Theoretical Framework

Prior research on the vertical effects of competition in financial intermediation often draws on insights from standard industrial organization theory (e.g. Cetorelli and Gambera (2001); Black and Strahan (2002); di Patti and Dell'Ariccia (2004)).

Traditional industrial organization models predict that market concentration influences prices and quantities through its effect on market power. In competitive markets, a larger number of suppliers reduces markups and increases the quantity of goods or services supplied. Applied to financial intermediation, this framework implies that greater competition among intermediaries should lower the cost of capital and increase the availability of external financing to firms. Under this view, a more competitive financial sector promotes entrepreneurial activity by expanding the supply of credit and reducing financing costs.

However, financial intermediation differs from standard product markets because it is characterized by substantial informational frictions. Financing young firms typically involves severe information asymmetries between entrepreneurs and investors, as startups often lack collateral, operating histories, and verifiable performance metrics (e.g. Jensen and Meckling (1976); Stiglitz and Weiss (1981); Berger and Udell (1998); Cosh et al. (2009)). These characteristics make screening, monitoring, and evaluating entrepreneurial ventures costly and uncertain.

Petersen and Rajan (1995) propose a framework in which banks form lending relationships with entrepreneurs in order to accumulate private information and mitigate these information asymmetries. In this setting, greater market power may facilitate relationship lending because banks have stronger incentives and greater capacity to sustain long-term relationships with borrowers. In particular, lenders may initially subsidize young firms and recoup these costs later through future rents once firms become more established (Petersen and Rajan 1995). By contrast, in highly competitive credit markets lenders may struggle to appropriate these future rents, weakening their incentives to finance early-stage or informationally opaque firms (Petersen and Rajan 1995).

Therefore, when information frictions are present, the relationship between market structure and financing outcomes becomes theoretically ambiguous. On the one hand, as suggested by traditional industrial organization theory, increased competition may improve access to finance by lowering prices and expanding the supply of funding. On the other hand, and under Petersen and Rajan (1995)'s lending relationship framework, greater market power may encourage intermediaries to invest in costly information production and monitoring, as the ability to capture future rents can justify these upfront investments. In such settings, market concentration may improve financing outcomes for informationally opaque firms.

To the best of our knowledge, no comparable theoretical framework exists for the venture capital industry that formally resolves this trade-off. Nonetheless, existing theory and empirical evidence suggest that similar mechanisms may operate in venture capital markets.

Venture capital represents a specialized segment of financial intermediation characterized by high levels of information asymmetry and intensive investor involvement in portfolio firms. Venture capitalists engage in extensive screening, staged financing, and governance activities to mitigate agency problems and improve firm outcomes (e.g. Gompers (1995); Kaplan et al (2001); Winton and Yerramilli (2008)). These activities require specialized expertise, extensive networks, and substantial fixed costs.

As a result, greater market power could lead to the emergence of larger and more established venture capital firms that benefit from stronger networks, better deal flow, and superior screening capabilities (e.g. Hochberg et al (2007)). These advantages may enhance investors' ability to identify and support high-potential startups, paralleling the benefits associated with bank market power in the relationship-lending framework.

At the same time, higher concentration may reduce competitive pressure in the market for entrepreneurial finance, potentially limiting funding availability or increasing investors' bargaining power *vis-à-vis* entrepreneurs (e.g. Hsu (2004)). This mechanism is consistent with standard industrial organization models, which predict that greater market power reduces the quantity of financing supplied. In addition, higher venture capital concentration may narrow the range of technologies and founders that receive funding, potentially leading to weaker social and economic outcomes (Lerner and Nanda 2020; The Purposeful Company 2025; Bonelli 2025).

Taken together, these mechanisms imply that the relationship between venture capital concentration and startup financing is theoretically ambiguous. Greater concentration may improve the efficiency of screening and monitoring by enabling specialized investors to develop expertise and networks. However, it may also reduce the supply of entrepreneurial finance if fewer investors compete to fund startups. In such settings, a small group of investors may act as gatekeepers within the entrepreneurial ecosystem, influencing which firms receive financing and which innovations are ultimately developed. These opposing mechanisms imply that the effect of market concentration in venture capital markets cannot be determined *a priori*.

2.2. Hypothesis Development

A clear burning question arises from the previous discussion: *what are the effects of VC concentration?*

The theoretical considerations outlined above suggest that the market structure of VC markets may influence startup financing and entrepreneurial activity. In particular, if an exogenous shock increases (decreases) concentration among VC investors, competition in the market for entrepreneurial finance may decrease (increase). Greater (lower) competition could expand (contract) the availability of capital, improve (deteriorate) financing terms, and lower (increase) barriers to entry for new ventures.

This reasoning leads to the following testable hypotheses.

Hypothesis 1.

A increase (decrease) in venture capital market concentration decreases (increases) the availability of funding for startups.

Greater (lower) competition among investors may also stimulate (reduce) entrepreneurial activity by improving (worsening) access to external finance. If more (less) investors compete to fund startups, the expected probability of obtaining financing increases (decreases), which may encourage more (less) entrepreneurs to enter the market.

Hypothesis 2.

A increase (decrease) in venture capital market concentration decreases (increases) startup entry.

Beyond financing and firm creation, VC market structure may also influence innovation outcomes. Prior research shows that venture capital investment tends to concentrate in a relatively narrow set of sectors and technologies (Lerner and Nanda 2020). If increased (reduced) concentration expands (contracts) the set of investors participating in the market, it may broaden (narrow) the range of technologies that receive funding.

Hypothesis 3.

A increase (decrease) in venture capital market concentration decreases (increases) innovation.

Finally, the composition of funded entrepreneurs may also depend on the structure of VC markets. Venture capital networks often exhibit strong homophily, which can limit funding access for underrepresented founders (Ewens and Townsend 2020; Calder-Wang and Gompers 2021). If a small number of investors dominate the market, such patterns may become amplified. By contrast, a more competitive venture capital market with a larger set of investors may increase the diversity of funded founders.

Hypothesis 4.

A increase (decrease) in venture capital market concentration decreases (increases) founder diversity.

3. THE ALTERNATIVE INVESTMENT FUND MANAGERS DIRECTIVE

Ideally, one could study the causal effects of venture capital market concentration through a controlled experiment in which otherwise similar startups are randomly assigned to markets with different levels of concentration. In practice, however, conducting such an experiment is infeasible.

Instead, and consistent with the "credibility revolution" in empirical economics (Angrist and Pischke 2010), we exploit the adoption of the Alternative Investment Fund Managers Directive (AIFMD; Directive 2011/61/EU) in Europe as a natural experiment to study how changes in venture capital market concentration affect startup outcomes. To the best of our knowledge, this paper is the first to use the AIFMD as a source of quasi-experimental variation in the venture capital industry.

The AIFMD formed part of the broader regulatory overhaul introduced in the aftermath of the Global Financial Crisis. Its primary objective was to increase transparency, strengthen investor protection, and enhance regulatory oversight of alternative investment funds. In this sense, the AIFMD can be viewed as broadly analogous to the Dodd–Frank Act in the United States, which similarly increased scrutiny of the private equity and financial intermediation sectors following the financial crisis (White, Mary Jo 2015). As documented in prior work, Dodd–Frank significantly affected the US venture capital ecosystem (Tillman 2012; Chen and Ewens 2025). Similarly, as reported by European Commission et al. (2025), the AIFMD had substantial effects on the European private equity and venture capital industries.

Prior to the AIFMD, many European jurisdictions applied relatively "light-touch" regulation to institutional investment funds and their managers (Latham & Watkins 2018). The Directive fundamentally altered this environment by introducing a comprehensive regulatory framework for Alternative Investment Fund Managers (AIFMs) operating within the European Union.

The Directive introduced several new requirements for fund managers, including minimum capital standards, stricter rules governing conflicts of interest, risk and liquidity management requirements, independent asset valuation procedures, and enhanced disclosure obligations (Clifford Chance 2012; Qadir 2016). It also imposed extensive transparency and reporting requirements designed to reduce information asymmetries and strengthen regulatory oversight. For example, AIFMs must produce annual reports for each fund they manage or market in the EU and disclose detailed information regarding their investment activities, principal exposures, and portfolio concentration (Clifford Chance 2012; Qadir 2016).

Beyond strengthening transparency and oversight, the AIFMD also aimed to harmonize the regulatory framework governing how alternative investment funds are managed and marketed across EU Member States. A central component of the Directive is the so-called “passport regime.” Under this system, an authorized AIFM can market and manage funds across all EU Member States after obtaining a single authorization from its home-country regulator. To use the passport, an AIFM must notify its home regulator of the Member States in which it intends to operate and provide documentation such as the fund’s governing rules and investor disclosures. The home regulator must then notify the relevant host regulators within 20 days. Once this notification is confirmed, the manager may begin marketing in those jurisdictions (Clifford Chance 2012).

Importantly, the AIFMD regulates fund managers rather than the funds themselves. The Directive applies to AIFMs engaged in the management or marketing of alternative investment funds within the EU (Clifford Chance 2012; Qadir 2016). Its scope is also limited to professional investors, defined as investors classified as professional clients under Annex II of Directive 2004/39/EC (Qadir 2016). In practice, private equity and venture capital funds are exclusively financed by such professional investors (European Commission et al. 2025).

The Directive applies fully to AIFMs managing portfolios with assets under management exceeding €100 million when leverage is used, or €500 million when leverage is not used and the fund has a lock-up period of at least five years (Clifford Chance 2012). Managers below these thresholds face lighter regulatory obligations and must only register with their home regulator and provide limited information regarding their activities (Clifford Chance 2012).

The AIFMD also restricts the ability of non-EU fund managers to market funds in the European Union. Non-EU managers cannot access the EU passport and must instead rely on the National Private Placement Regime (NPPR), which requires separate authorization in each Member State in which they intend to operate (Qadir 2016). Unlike the passport system, the NPPR does not provide a single EU-wide approval process. Managers must navigate multiple national registration procedures and comply with transparency requirements in each jurisdiction, including regulatory reporting and investor disclosure obligations (Qadir 2016).

While the Directive aimed to harmonize regulation across Europe, in practice it introduced substantial compliance costs and administrative complexity. Evidence from the European Commission’s 2025 evaluation of the AIFMD indicates that the costs associated with full regulatory compliance can exceed €1 million, particularly due to legal, consulting, and regulatory advisory fees incurred during the authorization process (European Commission et al. 2025). Moreover, ongoing compliance costs - including custodian bank fees, governance requirements, and additional administrative personnel - can represent approximately 20 percent of a fund’s operational expenses (European Commission et al. 2025).

These costs disproportionately affect smaller and younger fund managers. Funds at their early-stage often operate with relatively low profitability due to the inherent timing mismatch between upfront operational expenditures and the back-loaded structure of management fees and carried interest (European Commission et al. 2025). At the same time, smaller funds face higher cost ratios because fixed overhead costs weigh more heavily on a limited asset base (European Commission et al. 2025). Funds with less than €100 million in assets under management frequently operate below cost-recovery levels during the early years of their lifecycle (European Commission et al. 2025). In contrast, larger managers benefit from economies of scale in back-office operations and regulatory compliance (European Commission et al. 2025).

These regulatory burdens may also influence strategic decisions within the industry. For instance, managers approaching the regulatory thresholds may choose not to scale up their funds to avoid triggering full AIFMD compliance (European Commission et al. 2025). Instead, they may deliberately remain below the threshold or spin out new investment vehicles (European Commission et al. 2025). Such behavior can contribute to fragmentation in the venture capital market (European Commission et al. 2025).

The Directive may also discourage consolidation through mergers and acquisitions. Notably, the AIFMD does not provide specific provisions governing the regulatory treatment of M&A transactions involving AIFMs (European Commission et al. 2025). Instead, oversight remains largely delegated to national regulators, which apply domestic supervisory rules on a case-by-case basis (European Commission et al. 2025). This decentralized framework can generate legal uncertainty and regulatory delays, particularly for cross-border transactions (European Commission et al. 2025).

Industry interviews further highlight practical challenges associated with consolidation, including the integration of funds with different legal structures, divergent valuation methodologies, and heterogeneous groups of limited partners. These complexities can reduce the expected synergies from consolidation and weaken incentives for mergers among fund managers (European Commission et al. 2025).

Although the passport regime was intended to facilitate cross-border fundraising and investment, its benefits appear to have been limited in practice. The Directive left several key regulatory concepts undefined, allowing national regulators significant discretion in interpretation. Member States also retained substantial flexibility in how they transposed the Directive into domestic law. As a result, the AIFMD has generated a fragmented regulatory landscape across Europe. Some jurisdictions imposed additional compliance requirements - often referred to as regulatory “gold-plating” - which further increased costs and administrative complexity for fund managers (European Commission et al. 2025).

The Directive was published in the Official Journal of the European Union on July 1, 2011, and entered into force on July 21, 2011 (Clifford Chance 2012). However, Member States were given until July 22, 2013 to transpose the Directive into national law. In practice, several countries implemented the rules later, resulting in a staggered implementation process that extended from 2013 to 2016. Table 1, in the Appendix, reports the implementation dates across Member States and identifies the national legislation used to transpose the Directive.

The staggered implementation of EU directives provides a particularly attractive quasi-experimental setting for empirical research. Because Member States adopt EU directives at different times and retain discretion in implementation and enforcement, these policies generate cross-country and temporal variation that can be exploited for identification (e.g. Christensen et al. (2016); Ortiz et al. (2023)).

Finally, the discretionary nature of national implementation may mitigate concerns regarding anticipation effects. Political negotiations, legal uncertainty, and regulatory discretion created substantial ambiguity regarding the timing, scope, and strictness of the final rules. Contemporary academic and policy discussions link such uncertainty to regulatory “gold-plating” and heterogeneous implementation across Member States (e.g. Squintani (2019); ecoDa (2024); Thomsen (2025)).

4. DATA

We collect our investment-level data from LSEG Workspace. LSEG Workspace, and its previous iterations (Refinitiv, Thomson ONE, and VentureXpert), has been extensively used in the venture capital literature (e.g. Huang et al. (2021); Hellmann et al. (2021); Denes et al. (2023); Chen and Hshieh (2024)).

Our sample covers seed, early-stage, and late-stage venture capital investments. We focus on the 2008–2018 period, which provides a symmetric window of five years before and five years after the treatment year.

We first rank countries by the number of venture capital funding rounds. This ranking maximizes data coverage while ensuring cross-country comparability in the vibrancy and depth of venture capital ecosystems.

We then focus on the five largest European economies that implemented the AIFMD in 2013 and define our control group following related literature. Ortiz et al. (2023) examine the impact of Directive 2003/51/EC on M&A involving private targets, exploiting staggered adoption across countries and using the United States as a counterfactual. More broadly, cross-country studies in the finance literature that exploit legal reforms as natural experiments often benchmark European countries against a broad and heterogeneous set of non-European economies (e.g. Lel and Miller (2015); Fauver et al. (2017)).

Our approach lies between these two strategies. Following Ortiz et al. (2023), whose empirical setting shares similarities to ours, we include the United States as a counterfactual. We further expand the control group to other large and active venture capital markets among the world's largest economies by deal activity - namely China, Canada, India, and Japan - thereby ensuring a minimum degree of comparability in terms of market vibrancy and depth.

Our sample is restricted to a relatively small number of countries due to technical difficulties with LSEG Workspace that temporarily limited our access to data. Table 2 and Table 3 in the Appendix provide an overview of the sample.

In cases where investment is done in syndicates, we borrow from the literature on syndicated lending (e.g. Keil and Müller (2020); Saidi and Streitz (2021)) and we focus our analysis on the lead venture capital investor. An attractive feature from LSEG Workspace is the inclusion of a flag for the lead investor, when a startup has more than one investor in a one round.

5. THE IMPACT OF THE AIFMD ON VENTURE CAPITAL CONCENTRATION

5.1. METHODOLOGY

As a first stage of our study, we focus on determining if the AIFMD had an impact on the concentration levels of venture capital markets in Europe.

We conduct our analysis at the country-industry level. Our outcome variable is the Herfindahl–Hirschman Index (HHI), a standard measure of market concentration that has been used in the recent finance literature to study concentration in financial intermediation (e.g. Saidi and Streitz (2021)).

The HHI is bounded between zero and one, which may bias ordinary least squares (OLS) estimates, as documented by Papke and Wooldridge (1996; 2008). Wallis (1987) proposes the logistic transformation for Time Series analysis of bounded outcome variables,

$$\ln\left(\frac{y_t}{1-y_t}\right)$$

A limitation of this transformation is that it produces undefined values when the HHI equals one, leading to the exclusion of affected observations and a further reduction in sample size. While earlier practice often addressed this issue by adding an arbitrary constant to the dependent variable, recent work by Cohn et al. (2022) strongly discourages this approach, as it generates economically meaningless elasticities and distorts interpretation.

Accordingly, we adopt the transformation proposed by Wallis (1987) without *ad hoc* adjustments. To ensure that our results are not driven by the choice of functional form, we also propose an analysis using the natural logarithm of the HHI, $\ln(y_t)$.

We restrict our sample to country-industry pairs with complete data over the full analysis period. By focusing on balanced country–industry panels, we ensure a high level of data quality and internal consistency, mitigating concerns related to missing observations, differential sample composition over time, and measurement error. Table 4 lists treated and control countries and corresponding industries.

Our baseline specification is as follows:

$$Y_{it} = \alpha + \beta . D_{it} + \mu_i + \delta_t + \epsilon_{it}$$

Where:

Y_{it} = Outcome variable.
D_{it} = $Treated_i \times Post_t$.
μ_i = unit fixed effects.
δ_t = time fixed effects.
β = Average Treatment Effects (TWFE estimator).

We abstain from using controls, as these can be problematic with the Two-Way Fixed Effects (TWFE) estimator (e.g. Caetano et al. (2024)).

Following recent finance literature (e.g. Ortiz et al. (2023); Campello et al. (2024); Bonelli (2025); Core (2025); Chen and Ewens (2025)) we assess the parallel-trends assumption using event study specifications.

$$Y_{it} = \alpha + \sum_{k \neq 1} \beta_k \mathbb{1}(t - T_i = k) + \mu_i + \delta_t + \epsilon_{it}$$

As per the section before, we assume the absence of anticipation effects as gold-plating mitigates concerns about anticipation effects.

5.2. RESULTS AND PARALELL-TRENDS

Our results are presented in Table 4 in the Appendix.

Using our alternative transformations of market concentration, we find consistent evidence that treatment reduces HHI. In a TWFE DID specification with log-transformed HHI, the estimated coefficient of −0.284 implies a proportional decline in concentration of approximately 24.7 percent relative to the counterfactual. Results are qualitatively similar when using a logit transformation of HHI, which accounts explicitly for the bounded nature of the index. The logit estimate of −0.342 corresponds to approximately a 29 percent reduction in the odds of concentration and translates into economically meaningful declines in HHI levels, particularly in more concentrated markets. Taken together, these results indicate a sizable and robust reduction in market concentration following treatment, independent of the specific functional form used. Both estimates are significant at 5 percent level.

These results are robust to the parallel-trends assumption, as showed in Table 5 and event study plots (Figure 1 and Figure 2) in the Appendix.

To further strengthen the robustness of our findings to the parallel-trends assumption, we follow Campello et al. (2024) and complement our baseline estimates with results from the Synthetic Difference-in-Differences estimator (SDID) proposed by Arkhangelsky et al. (2021).

The results remain statistically significant, with the log-transformed HHI yielding estimates that closely align with the baseline TWFE DID results. Some differences emerge when using the logit transformation of the HHI, reflecting the exclusion of additional units due to undefined values and the resulting reduction in sample size.

5.3. ROBUSTNESS CHECKS

We conduct extensive robustness checks. These are summarized in Table 6.

First, we examine alternative time windows in a series of robustness checks and find that our results remain statistically significant.

Second, we focus on the treatment group and conduct leave-one-out (LOO) and stepwise inclusion tests of additional European countries (the Netherlands, Denmark, and Ireland). Our findings are generally robust, the sole exception being when excluding the United Kingdom, in which case statistical significance is lost. This result is plausibly driven by the substantial reduction in the number of observations, as this specification yields the smallest sample size among all robustness checks. Importantly, across all specifications, the estimated effects and standard errors remain stable.

We also conduct LOO and stepwise inclusion tests (Australia, Singapore and Israel) focusing on the control group. Our results remain statistically significant. In addition, the estimated effects remain economically stable in magnitude, and standard errors are largely consistent.

6. THE IMPACT OF THE AIFMD AT THE STARTUP LEVEL

6.1. METHODOLOGY

In the second stage of our analysis, we focus on the effects at the startup level.

Aligned with our previous country–industry–level analysis, we focus on startups founded in the United Kingdom, France, Germany, Sweden, and Switzerland. We restrict the sample to firms seeking first-round financing, consistent with prior literature (Chen and Ewens 2025) and with the structure of our data, as many startups no longer appear in LSEG Workspace after their initial funding round. Accordingly, the dataset is best characterized as a repeated cross-section.

In addition, we restrict our sample to startups at the Seed and Early Stage, to ensure consistency across funding requirements. For example, the Later-Stage group includes firms at different stages of their life and at different several rounds of funding with, also, very heterogenous funding needs.

Our control group consists of startups founded in the United States. We do so, again, to ensure cross-country heterogeneity in funding scale. For example, startups in countries such as China or India - where average wages and operating costs (e.g. rent) are lower - likely require smaller funding rounds than comparable startups in higher-cost European economies such as France, Germany, or Switzerland. This heterogeneity inflates dispersion in monetary outcomes, reducing statistical precision and potentially masking economically meaningful effects. Thus, we decided to restrict our counterfactual to startups founded in the United States.

Our outcome variables are the total equity raised by the startup in the round (Round_Equity), the number of investors participating in the round (No_Investors), and the average investment per investor (Avg_Investment). We apply a logarithmic transformation to each outcome variable. In addition, we winsorize the outcome variables at the 1st and 99th percentiles.

Our baseline specification is the same as in the previous section, and we also test for the parallel-trends assumption using event study specifications.

Rather than relying on control variables, we match startups using Propensity Score Matching (PSM). Matching is performed on average equity funding (“ind_avg_equity_pre”) and average investor per deal (“ind_avg_investors_pre”). These are calculated at the industry level and pre-treatment (thus, they are time-invariant). In addition, we match on startup age at funding round (“age_at_funding”).

Consistent with our prior country–industry–level analysis, we focus on the 2008–2018 period, which provides a symmetric five-year window before and after the treatment year. We additionally assume the absence of anticipation effects.

6.2. RESULTS AND PARALELL-TRENDS

Our results are presented in Table 8 in the Appendix.

Our results show that, following the implementation of the AIFMD and the associated increase in market fragmentation, venture capital financing outcomes for treated startups declined relative to the U.S. control group.

Specifically, the coefficient on the treatment indicator in the log equity specification is −0.323, implying that startups in treated European markets raised approximately 28 percent less capital per funding round relative to comparable U.S. startups after the policy change. Similarly, the number of investors participating in a round declines by about 7 percent. The average investment per investor also falls significantly, corresponding to a reduction of roughly 22 percent.

Taken together, these findings suggest that the implementation of AIFMD is associated with a contraction in venture financing at the startup level, affecting both the size of funding rounds and the composition of investor syndicates. The decline in total funding appears to arise from both fewer participating investors and smaller individual ticket sizes, consistent with the interpretation that regulatory compliance costs or barriers to cross-border fundraising may have reduced the supply of venture capital available to European startups in the post-AIFMD period.

These results are largely robust to the parallel-trends assumption, as shown in in the event-study plots reported in the Appendix (Figure 3, Figure 4, Figure 5). The only exception concerns the estimate for syndicate size (number of investors), for which the event-study evidence suggests the presence of slight pre-treatment differences, indicating that this coefficient may be subject to some bias.

6.3. ROBUSTNESS CHECKS

We conduct a series of robustness checks using alternative matching strategies. The results are reported in Table 9.

In a first set of tests, we retain the same set of matching variables but modify the matching procedure. Specifically, we re-estimate our baseline specification using Coarsened Exact Matching (CEM) and Entropy Balancing. In addition, we estimate the model using the full unmatched sample. Across these specifications, the results remain broadly consistent with the baseline estimates in terms of economic magnitude, standard errors, and statistical significance. The main exception concerns the average investment per investor, which loses statistical significance.

In a second set of robustness checks, we remove the variable average investor per deal from the set of matching variables and repeat all specifications (PSM, CEM, Entropy Balancing, and the unmatched estimation). The results remain largely unchanged, displaying similar economic magnitudes and levels of statistical significance. As before, the coefficient on average investment per investor becomes statistically insignificant.

Finally, we perform an additional set of tests excluding average equity funding from the matching variables while again estimating all specifications (PSM, CEM, Entropy Balancing, and the unmatched model). The results continue to be broadly consistent with the baseline findings. As in the previous exercises, the estimate for average investment per investor loses statistical significance, while the remaining outcome variables remain stable.

7. DISCUSSION, LIMITATIONS AND FUTURE EXTENSIONS

Overall, the evidence suggests that the AIFMD led to fragmentation in the European venture capital market. This disproves our hypothesis 1, in the sense that the AIFMD – acting as an exogenous shock - decreased venture capital market concentration and, consequently, to a reduction the availability of funding for startups.

At the startup level, this led to a decrease in total startup funding, syndicate size and average investment per investor.

As stated in a previous section, the AIFMD created incentives for managers to remain below regulatory thresholds, effectively constraining fund size. With smaller funds and less capital under management, the aggregate amount of capital available for startup financing would naturally decline, consistent with the observed reduction in funding – either at the aggregate and in terms of per investor ticket.

The contraction in syndicate size - assuming investors did not exit the European market - further suggests a shift in investment behavior. Faced with tighter capital constraints, investors likely became more selective, participating in fewer deals in order to allocate limited resources more carefully across portfolio companies.

We are currently working towards estimates that might help us confirm the remainder of our hypothesis.

Indeed, **this paper remains a work in progress**, and substantial scope remains for improvements and extensions that may strengthen its contribution to the literature. Unfortunately, technical problems related to LSEG Workspace significantly limited our access to data for an extended period, which delayed our progress. As a result, several pragmatic decisions were made in order to produce a viable - albeit imperfect - research paper within the required timeframe.

In terms of limitations:

1. Our HHI measure is computed using the Thomson Reuters Business Classification (TRBC) system. However, traditional industry classification systems are known to suffer

from several limitations, including their inability to capture product–market similarities across firms (e.g. Hoberg and Philips (2010; 2016)). In the spirit of Hoberg and Phillips (2016; 2025), we developed a bespoke startup taxonomy based on textual analysis and machine-learning techniques applied to firm descriptions. Due to the technical constraints discussed above, this taxonomy currently covers only a subset of firms in our dataset. In future work, we plan to expand its coverage to the full sample. Traditional classification systems such as TRBC and SIC will be used in robustness checks.

2. Investor names provided by LSEG Workspace are not standardized, meaning that the same investor may appear under multiple name variations. The finance literature has developed methods to address this issue (e.g. González-Uribe (2020); Brogaard et al. (2021)), typically involving extensive manual cleaning and validation procedures. Given the technical constraints described above, we were able to standardize investor names for only a subset of observations in our dataset. Extending this standardization process to the full sample is part of our planned future work.
3. Given the time required to address the technical challenges described above, we opted to implement a simpler non-staggered research design using the standard TWFE estimator. In future work, we plan to revisit this aspect of the empirical strategy and explore alternative approaches that may be more suitable given the structure of our data. In particular, we aim to implement a staggered design using difference-in-differences estimators such as those proposed by Callaway and Sant'Anna (2021), as well as the Extended Two-Way Fixed Effects (ETWFE) estimator proposed by Wooldridge (2023; 2025).
4. Implementing a staggered research design would also allow us to include additional treated countries in the analysis, including large economies such as Poland, Spain, and Italy. While our results are generally robust to the inclusion or exclusion of individual countries, expanding the set of treated economies would increase the external validity of our findings. Moreover, incorporating a broader set of countries within a staggered design may help establish a more refined and coherent counterfactual across both levels of analysis and strengthen the plausibility of the parallel-trends assumption, which may currently introduce some bias in parts of the startup-level analysis.
5. In future work, we also plan to explore additional empirical specifications to ensure that our results remain robust across alternative econometric approaches. One promising avenue is the implementation of a dosage-type research design, similar to that employed by Chen and Ewans (2025), which exploits cross-sectional variation in exposure to

regulatory changes. Another potential strategy involves the use of Bartik (1991)-type instruments (see also Goldsmith-Pinkham et al. (2020)) to capture differential exposure to the policy shock. Both approaches could leverage the threshold-based structure of the AIFMD.

6. The startup-level analysis could also benefit from an expanded set of robustness checks to ensure that the results are not driven by any specific modeling assumption or empirical specification.
7. At this stage, the empirical analysis focuses on testing a single hypothesis. In future work, we plan to extend the analysis to examine the remaining hypotheses. In particular, testing Hypotheses 3 and 4 will likely require integrating our dataset with additional sources of information, such as patent data and employee-level datasets (e.g. from Revelio Labs). This would allow us to explore more granular outcomes related to innovation and diversity within startups.
8. Related research examining policy shocks typically provides a comprehensive analysis of the policy's effects across the broader ecosystem. For example, Chen and Ewans (2025) provide a detailed discussion of the extent to which venture capital funds relied on bank financing and how the Volcker Rule affected subsequent fundraising activity. Similarly, Core (2025) offers an in-depth analysis of how bank concentration interacts with the policy mechanisms introduced by the Start-Up Italy Act. In future work, we aim to follow a similar approach by developing a more comprehensive analysis of the effects of the AIFMD on the European venture capital ecosystem.

Regarding potential extensions, the methodology underlying the startup taxonomy described above can be readily adapted to construct a technology-based taxonomy (see, for example, Arts et al., (2023)). This would allow us to classify startups based on their underlying technological domains rather than solely on their business models or product-market characteristics. Consequently, in addition to analyzing concentration from a business or product-market perspective, this approach would enable us to examine whether the AIFMD has affected the concentration of investment across technological domains. Such an analysis could provide additional insights into whether regulatory changes influence not only the allocation of venture capital across firms, but also the direction of technological development within the startup ecosystem. This could have substantial policy implications considering current geopolitical concerns surrounding technological sovereignty.

8. CONCLUSION

This paper studies how the market structure of venture capital markets influences startup financing outcomes. While venture capital plays a central role in financing innovation and supporting high-growth entrepreneurship, the effects of concentration among venture capital investors remain poorly understood.

To address this question, we exploit the implementation of the Alternative Investment Fund Managers Directive (AIFMD) as a quasi-natural experiment that altered the regulatory environment of venture capital markets in Europe. Using investment-level data from LSEG Workspace covering the period 2008–2018, we first analyze how the directive affected concentration levels in venture capital markets and then examine how these changes influenced startup-level financing outcomes.

Our findings indicate that the AIFMD significantly reduced venture capital market concentration, leading to greater fragmentation among investors. This increase in fragmentation is associated with weaker financing outcomes for startups. Following the implementation of the directive, startups in treated European markets raise less capital per funding round, attract fewer investors, and receive smaller investments from each participating investor relative to comparable startups in the United States.

These results suggest that the relationship between competition and financing outcomes in venture capital markets differs from the predictions of standard industrial organization models. In settings characterized by substantial information asymmetries and high fixed costs of screening and monitoring, greater fragmentation among investors may reduce the effective supply of capital available to startups.

Our findings therefore highlight the importance of market structure in venture capital ecosystems and suggest that regulatory interventions can have unintended consequences for entrepreneurial finance. Policies that increase compliance costs or discourage the scaling of venture capital funds may reduce the availability of capital for startups, potentially slowing innovation and economic growth.

Overall, this paper contributes to a growing body of research examining the industrial organization of financial intermediation and highlights the critical role of venture capital market structure in shaping entrepreneurial ecosystems.

9. REFERENCES

Angrist, J.D. and Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24(2), pp.3–30.

Arkhangelsky, D. et al. (2021). Synthetic Difference-in-Differences. *American Economic Review*, 111(12), pp.4088–4118.

Arts, S., Cassiman, B. and Hou, J. (2023). Position and Differentiation of Firms in Technology Space. *Management Science*, 69(12), pp.7253–7265.

Bartik, T.J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute. [online]. Available from: http://www.jstor.org.manchester.idm.oclc.org/stable/j.ctvh4zh1q [Accessed March 10, 2026].

Beck, T., Demirgüç-Kunt, A. and Maksimovic, V. (2004). Bank Competition and Access to Finance: International Evidence. *Journal of Money, Credit and Banking*, 36(3), pp.627–648.

Bellia, M. et al. (2025). Market Liquidity and Competition Among Designated Market Makers. *Management Science (INFORMS)*, 71(1), pp.184–201.

Black, S.E. and Strahan, P.E. (2002). Entrepreneurship and Bank Credit Availability. *The Journal of Finance*, 57(6), pp.2807–2833.

Bonelli, M. (2025). Data-Driven Investors. *The Review of Financial Studies*, p.hhaf078.

Brav, A. and Gompers, P. (1997). Myth or Reality? The Long-Run Underperformance of Initial Public Offerings: Evidence from Venture and Nonventure Capital-Backed Companies. *The Journal of Finance*, 52(5), pp.1791–1821.

Brogaard, J., Denes, M. and Duchin, R. (2021). Political Influence and the Renegotiation of Government Contracts. *The Review of Financial Studies*, 34(6), pp.3095–3137.

Caetano, C. et al. (2024). Difference in Differences with Time-Varying Covariates. [online]. Available from: http://arxiv.org/abs/2202.02903 [Accessed January 19, 2026].

Calder-Wang, S. and Gompers, P.A. (2021). And the children shall lead: Gender diversity and performance in venture capital. *Journal of Financial Economics*, 142(1), pp.1–22.

Callaway, B. and Sant'Anna, P.H.C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), pp.200–230.

Campello, M., Kankanhalli, G. and Kim, H. (2024). Delayed creative destruction: How uncertainty shapes corporate assets. *Journal of Financial Economics*, 153, p.103786.

Cetorelli, N. (2004). Real Effects of Bank Competition. *Journal of Money, Credit and Banking*, 36(3), pp.543–558.

Cetorelli, N. and Gambera, M. (2001). Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data. *The Journal of Finance*, 56(2), pp.617–648.

Cetorelli, N. and Strahan, P.E. (2006). Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets. *The Journal of Finance*, 61(1), pp.437–461.

Chemmanur, T.J., Loutskina, E. and Tian, X. (2014). Corporate Venture Capital, Value Creation, and Innovation. *The Review of Financial Studies*, 27(8), pp.2434–2473.

Chen, J. and Ewens, M. (2025). Venture Capital and Startup Agglomeration. *The Journal of Finance*, 80(4), pp.2153–2198.

Chen, J. and Hshieh, S. (2024). The labor effects of R&D tax incentives: evidence from VC-backed startups. *Review of Finance*, 28(5), pp.1451–1482.

Christensen, H.B., Hail, L. and Leuz, C. (2016). Capital-Market Effects of Securities Regulation: Prior Conditions, Implementation, and Enforcement. *The Review of Financial Studies*, 29(11), pp.2885–2924.

Clifford Chance. (2012). *The Alternative Investment Fund Managers Directive – An Overview*. London, United Kingdom: Clifford Chance. [online]. Available from: https://www.cliffordchance.com/content/dam/cliffordchance/briefings/2012/06/the-alternative-investment-fund-managers-directive-an-overview.pdf.

Cohn, J.B., Liu, Z. and Wardlaw, M.I. (2022). Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2), pp.529–551.

Core, F. (2025). Bank market power and incentives for firm creation in innovative industries. *Review of Finance*, p.rfaf045.

Cosh, A., Cumming, D. and Hughes, A. (2009). Outside Enterpreneurial Capital. *The Economic Journal*, 119(540), pp.1494–1533.

Denes, M. et al. (2023). Investor Tax Credits and Entrepreneurship: Evidence from U.S. States. *The Journal of Finance*, 78(5), pp.2621–2671.

ecoDa. (2024). European Commission Corporate Sustainability Due Diligence proposal risks being counterproductive. [online]. Available from: https://ecoda.eu/wp-content/uploads/2019/08/20220523-ecoDa-Feedback-on-CSDD-Final-1.pdf.

European Commission et al. (2025). *Study of barriers to, and drivers of, the scaling-up of funds investing in innovative and growth companies*. Publications Office of the European Union. [online]. Available from: https://data.europa.eu/doi/10.2874/0565970 [Accessed March 2, 2026].

Ewens, M. and Townsend, R.R. (2020). Are early stage investors biased against women? *Journal of Financial Economics*, 135(3), pp.653–677.

Fauver, L. et al. (2017). Board reforms and firm value: Worldwide evidence. *Journal of Financial Economics*, 125(1), pp.120–142.

Feldman, D., Saxena, K. and Xu, J. (2020). Is the active fund management industry concentrated enough? *Journal of Financial Economics*, 136(1), pp.23–43.

Francis, B.B. et al. (2021). Do activist hedge funds target female CEOs? The role of CEO gender in hedge fund activism. *Journal of Financial Economics*, 141(1), pp.372–393.

Giambona, E., Kumar, A. and Phillips, G.M. (2025). Hedging, Contract Enforceability, and Competition I. Goldstein, ed. *The Review of Financial Studies*, 38(7), pp.2034–2087.

Goldsmith-Pinkham, P., Sorkin, I. and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8), pp.2586–2624.

Gompers, P.A. (1995). Optimal Investment, Monitoring, and the Staging of Venture Capital. *The Journal of Finance*, 50(5), pp.1461–1489.

González-Uribe, J. (2020). Exchanges of innovation resources inside venture capital portfolios. *Journal of Financial Economics*, 135(1), pp.144–168.

Hellmann, T., Schure, P. and Vo, D.H. (2021). Angels and venture capitalists: Substitutes or complements? *Journal of Financial Economics*, 141(2), pp.454–478.

Hoberg, G., Kumar, N. and Prabhala, N. (2018). Mutual Fund Competition, Managerial Skill, and Alpha Persistence. *The Review of Financial Studies*, 31(5), pp.1896–1929.

Hoberg, G. and Phillips, G. (2010). Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis. *The Review of Financial Studies*, 23(10), pp.3773–3811.

Hoberg, G. and Phillips, G. (2016). Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5), pp.1423–1465.

Hoberg, G. and Phillips, G.M. (2025). Scope, Scale, and Concentration: The 21st-Century Firm. *The Journal of Finance*, 80(1), pp.415–466.

Hochberg, Y.V., Ljungqvist, A. and Lu, Y. (2007). Whom You Know Matters: Venture Capital Networks and Investment Performance. *The Journal of Finance*, 62(1), pp.251–301.

Hsu, D.H. (2004). What Do Entrepreneurs Pay for Venture Capital Affiliation? *The Journal of Finance*, 59(4), pp.1805–1844.

Huang, J. and Kisgen, D.J. (2013). Gender and corporate finance: Are male executives overconfident relative to female executives? *Journal of Financial Economics*, 108(3), pp.822–839.

Huang, S. et al. (2021). Public Market Players in the Private World: Implications for the Going-Public Process. *The Review of Financial Studies*, 34(5), pp.2411–2447.

Inderst, R. and Mueller, H.M. (2009). Early-stage financing and firm growth in new industries. *Journal of Financial Economics*, 93(2), pp.276–291.

Jensen, M.C. and Meckling, W.H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), pp.305–360.

Kaplan, S.N. and Strömberg, P. (2001). Venture Capitalists as Principals: Contracting, Screening, and Monitoring. *The American Economic Review*, 91(2), pp.426–430.

Keil, J. and Müller, K. (2020). Bank Branching Deregulation and the Syndicated Loan Market. *Journal of Financial and Quantitative Analysis*, 55(4), pp.1269–1303.

Kerr, W.R., Lerner, J. and Schoar, A. (2014). The Consequences of Entrepreneurial Finance: Evidence from Angel Financings. *The Review of Financial Studies*, 27(1), pp.20–55.

Kerr, W.R. and Nanda, R. (2009). Democratizing entry: Banking deregulations, financing constraints, and entrepreneurship. *Journal of Financial Economics*, 94(1), pp.124–149.

Khorana, A. and Servaes, H. (2012). What Drives Market Share in the Mutual Fund Industry? *Review of Finance*, 16(1), pp.81–113.

Kostovetsky, L. and Warner, J.B. (2020). Measuring Innovation and Product Differentiation: Evidence from Mutual Funds. *The Journal of Finance*, 75(2), pp.779–823.

Latham & Watkins. (2018). *Alternative Investment Fund Managers Directive (AIFMD): An Overview*. London, United Kingdom: Latham & Watkins. [online]. Available from: https://www.lw.com/admin/Upload/Documents/AIFMD-QSG.v2.pdf.

Lel, U. and Miller, D.P. (2015). Does Takeover Activity Cause Managerial Discipline? Evidence from International M&A Laws. *The Review of Financial Studies*, 28(6), pp.1588–1622.

Lerner, J. (1994). Venture capitalists and the decision to go public. *Journal of Financial Economics*, 35(3), pp.293–316.

Lerner, J. and Nanda, R. (2020). Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn. *Journal of Economic Perspectives*, 34(3), pp.237–61.

Lindsey, L. (2008). Blurring Firm Boundaries: The Role of Venture Capital in Strategic Alliances. *The Journal of Finance*, 63(3), pp.1137–1168.

Loseto, Marco and Mainardi, Federico. (2024). Market Power, Fund Proliferation, and Asset Prices. *Working Paper*. [online]. Available from: https://share.google/VhDkJDWLn2jq7MMOU.

Malenko, Andrey, Malenko, Nadya, and Tsoy, Anton. (2026). Fund Proliferation, Decentralization, and Shareholder Power. *Working Paper*. [online]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6083066.

Massa, M. (2003). How do family strategies affect fund performance? When performance-maximization is not the only game in town. *Journal of Financial Economics*, 67(2), pp.249–304.

Megginson, W.L. and Weiss, K.A. (1991). Venture Capitalist Certification in Initial Public Offerings. *The Journal of Finance*, 46(3), pp.879–903.

N. Berger, A. and F. Udell, G. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance*, 22(6), pp.613–673.

Nahata, R. (2008). Venture capital reputation and investment performance. *Journal of Financial Economics*, 90(2), pp.127–151.

Ortiz, M. et al. (2023). Mandatory Financial Disclosure and M&A Activity. *The Review of Financial Studies*, 36(12), pp.4788–4823.

Papke, L.E. and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6), pp.619–632.

Papke, L.E. and Wooldridge, J.M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, 145(1), pp.121–133.

di Patti, E.B. and Dell'Ariccia, G. (2004). Bank Competition and Firm Creation. *Journal of Money, Credit and Banking*, 36(2), pp.225–251.

Petersen, M.A. and Rajan, R.G. (1995). The Effect of Credit Market Competition on Lending Relationships*. *The Quarterly Journal of Economics*, 110(2), pp.407–443.

Puri, M. and Zarutskie, R. (2012). On the Life Cycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms. *The Journal of Finance*, 67(6), pp.2247–2293.

Qadir, Z. (2016). Interpreting Vague Provisions of the Alternative Investment Fund Managers Directive Through a Systemic Cost-Benefit Analysis. *Chicago Journal of International Law*, 17(1).

Rice, T. and Strahan, P.E. (2010). Does Credit Competition Affect Small-Firm Finance? *The Journal of Finance*, 65(3), pp.861–889.

Saidi, F. and Streitz, D. (2021). Bank Concentration and Product Market Competition. *The Review of Financial Studies*, 34(10), pp.4999–5035.

Sørensen, M. (2007). How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *The Journal of Finance*, 62(6), pp.2725–2762.

Squintani, L. ed. (2019). Gold-Plating: A Misleading Overarching Concept. In *Beyond Minimum Harmonisation: Gold-Plating and Green-Plating of European Environmental Law*. Cambridge Studies in European Law and Policy. Cambridge: Cambridge University Press, pp. 13–71. [online]. Available from: https://www.cambridge.org/core/books/beyond-minimum-harmonisation/goldplating-a-misleading-overarching-concept/9320B6ECB7D6D5528CBFAD5FA6C6FE13 [Accessed January 19, 2026].

Stiglitz, J.E. and Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review*, 71(3), pp.393–410.

Sun, Y. (2021). Index Fund Entry and Financial Product Market Competition. *Management Science (INFORMS)*, 67(1), pp.500–523.

The Purposeful Company. (2025). *Venture Capital as 21st Century Growth Catalyst*. London, United Kingdom: The Purposeful Company. [online]. Available from: https://thepurposefulcompany.org/wp-content/uploads/2025/05/The-Growth-Trilogy-Paper-Two.pdf.

Thomsen, I.E. (2025). *Between Compliance and Control*. Thesis. Department of Political Science: Lund University. [online]. Available from: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9189251&fileOId=9189885.

Tillman, J.A. (2012). Beyond The Crisis: Dodd-Frank and Private Equity. *New York University Law Review*, 87(5), pp.1602–1640.

Wahal, S. and Wang, A. (Yan). (2011). Competition among mutual funds. *Journal of Financial Economics*, 99(1), pp.40–59.

Wallis, K.F. (1987). Time Series Analysis of Bounded Economic Variables. *Journal of Time Series Analysis*, 8(1), pp.115–123.

White, Mary Jo. (2015). Five Years On: Regulation of Private Fund Advisers After Dodd-Frank. *Keynote Address at the Managed Fund Association: "Five Years On: Regulation of Private Fund Advisers After Dodd-Frank"*. [online]. Available from: https://www.sec.gov/newsroom/speeches-statements/white-regulation-private-fund-advisers-after-dodd-frank [Accessed March 3, 2026].

Winton, A. and Yerramilli, V. (2008). Entrepreneurial finance: Banks versus venture capital. *Journal of Financial Economics*, 88(1), pp.51–79.

Wooldridge, J.M. (2023). Simple approaches to nonlinear difference-in-differences with panel data. *The Econometrics Journal*, 26(3), pp.C31–C66.

Wooldridge, J.M. (2025). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *Empirical Economics*, 69(5), pp.2545–2587.

APPENDIX A: TABLES AND GRAPHS

Table 1. AIFMD Adoption per European Union Member State, Norway and Switzerland.

This table lists the main AIFMD implementing act for each European Union Member State, together with the corresponding publication date and entry-into-force date. It also includes information for Norway and Switzerland, which likewise implemented the AIFMD framework. In addition, the table provides links to the relevant implementing acts for each jurisdiction.

Country	Main implementing law / act (includes link)	Date Published	Entry into Force	Year
Austria	Austrian AIFM Act	29/07/2013	22/07/2013	2013
Belgium	Act implementing AIFMD	17/06/2014	19/04/2014	2014
Bulgaria	Amendments to Collective Investment Schemes Act	20/12/2013	20/12/2013	2013
Croatia	Law on Alternative Investment Funds	08/02/2013	01/07/2013	2013
Cyprus	AIFM Law 56(I)/2013	05/07/2013	05/07/2013	2013
Czech Republic	Act No. 204/2013 on Investment Companies and Investment Funds	18/08/2013	18/08/2013	2013
Denmark	Danish AIFM law	12/06/2013	22/07/2013	2013
Estonia	Amendments to Investment Fund Act	11/07/2013	22/07/2013	2013
Finland	AIFMD Act 162/2014 + MoF decrees	07/03/2014	15/03/2014	2014
France	AIFMD transposition measures	27/07/2013	28/07/2013	2013
Germany	Kapitalanlagegesetzbuch (KAGB) + tax adaptation law	10/07/2013	22/07/2013	2013
Greece	Law 4209/2013	21/11/2013	21/11/2013	2013
Hungary	Act XVI of 2014 (The "AIFM Act")	24/02/2014	15/03/2014	2014
Ireland	AIFM Act (No. 257 of 2013) & Central Bank AIF Rulebook	16/07/2013	22/07/2013	2013
Italy	Legislative Decree 44/2014 + Consob/BoI regulations	25/03/2014	09/04/2014	2014
Latvia	AIFM Law "Alternatīvo ieguldījumu fondu un to pārvaldnieku likums"	24/07/2013	08/08/2013	2013
Lithuania	Law on Management Companies of Collective Investment Undertakings for Professional Investors	28/06/2013	01/07/2013	2013
Luxembourg	AIFM Law of 12 July 2013	12/07/2013	12/07/2013	2013
Malta	Amendments & Investment Services Rules under Investment Services Act	08/03/2013	22/07/2013	2013

Netherlands	Implementation legislation & amending bill	25/06/2013	22/07/2013	2013
Poland	Polish AIFM framework (amendments to financial markets acts)	04/05/2016	04/06/2016	2016
Portugal	National law transposing AIFMD	24/02/2015	24/03/2015	2015
Romania	Law 74/2015 on AIFMs	23/04/2015	23/05/2015	2015
Slovakia	Amendment to Collective Investment Act	19/07/2013	22/07/2013	2013
Slovenia	Law on Alternative Investment Fund Manager (ZUAIS)	08/05/2015	23/05/2015	2015
Spain	Law 22/2014 + Royal Decree	13/11/2014	14/11/2014	2014
Sweden	AIFM Act SFS 2013:561 + FSA rules	28/06/2013	22/07/2013	2013
United Kingdom	Alternative Investment Fund Managers Regulations 2013 (SI 2013/1773)	16/07/2013	22/07/2013	2013
Norway	Act on the Management of Alternative Investment Funds	20/06/2014	01/07/2014	2014
Switzerland	Loi fédérale sur les placements collectifs de capitaux	19/02/2013	01/03/2013	2013

Table 2. Descriptive Statistics per Year.

This table reports yearly descriptive statistics. Monetary values are expressed in millions of U.S. dollars.

Year	No. Rounds	No. Startups	No. Investors	No. Industries	No. Countries	Total Investiment	Average Round Size	Average Investor per Round
2008	2,323.00	2,152.00	645.00	104.00	54.00	16,677.59	7.18	1.47
2009	1,929.00	1,781.00	590.00	98.00	52.00	12,245.52	6.35	1.44
2010	2,306.00	2,140.00	645.00	111.00	57.00	15,842.11	6.87	1.43
2011	2,741.00	2,548.00	692.00	105.00	59.00	17,123.83	6.25	1.45
2012	2,628.00	2,460.00	712.00	102.00	57.00	15,577.79	5.93	1.49
2013	2,798.00	2,581.00	735.00	103.00	56.00	17,214.32	6.15	1.56
2014	3,136.00	2,885.00	761.00	102.00	58.00	29,360.31	9.36	1.62
2015	3,480.00	3,237.00	864.00	112.00	63.00	39,587.15	11.38	1.72
2016	3,113.00	2,935.00	880.00	103.00	61.00	35,849.93	11.52	1.76
2017	3,210.00	3,019.00	930.00	112.00	56.00	48,218.65	15.02	1.98
2018	3,733.00	3,530.00	1,039.00	112.00	68.00	70,209.62	18.81	2.09
Overall	**31,397.00**	**29,268.00**	**8,493.00**	**1,164.00**		**317,906.82**	**9.53**	**1.64**

Table 3. Descriptive Statistics per Country.

This table reports yearly descriptive statistics per country. Monetary values are expressed in millions of U.S. dollars.

Country	No. Rounds	No. Startups	No. Industries	No. Investors	Total Investment	Average Round Size	Average Investor per Round
Panel A: Treated Countries							
United Kingdom	1,492.00	1,199.00	101.00	520.00	12,791.02	8.57	1.52
France	1,196.00	1,062.00	96.00	275.00	5,486.58	4.59	1.24
Germany	473.00	390.00	66.00	235.00	3,997.22	8.45	1.84
Sweden	288.00	225.00	52.00	113.00	2,460.88	8.54	1.25
Switzerland	185.00	142.00	37.00	103.00	2,020.21	10.92	1.79
Overall	**3,634.00**	**3,018.00**	**352.00**	**1,246.00**	**26,755.92**	**8.22**	**1.53**
Panel B: Counterfactual Countries							
United States	18,091.00	11,929.00	122.00	1,226.00	173,776.47	9.61	1.84
China (Mainland)	3,154.00	2,561.00	110.00	140.00	64,264.07	20.38	1.51
Canada	1,302.00	905.00	89.00	383.00	6,454.81	4.96	1.46
India	950.00	767.00	91.00	127.00	11,792.88	12.41	1.38
Japan	464.00	431.00	74.00	87.00	1,323.03	2.85	1.16
Overall	**23,961.00**	**16,593.00**	**486.00**	**1,963.00**	**257,611.26**	**10.04**	**1.47**

Table 4. List of Treated and Control Country-Industry Pairs.

Country-Industries
Panel A: Treated
France-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
France-Industrials-Industrial & Commercial Services-Professional & Commercial Services-Business Support Services
France-Technology-Software & IT Services-Software & IT Services-IT Services & Consulting
France-Technology-Software & IT Services-Software & IT Services-Online Services
France-Technology-Software & IT Services-Software & IT Services-Software
Germany-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
Germany-Technology-Software & IT Services-Software & IT Services-Online Services
Germany-Technology-Software & IT Services-Software & IT Services-Software
Sweden-Technology-Software & IT Services-Software & IT Services-Software
Switzerland-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
United Kingdom-Healthcare-Healthcare Services & Equipment-Healthcare Equipment & Supplies-Advanced Medical Equipment & Technology
United Kingdom-Healthcare-Healthcare Services & Equipment-Healthcare Equipment & Supplies-Medical Equipment, Supplies & Distribution
United Kingdom-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
United Kingdom-Healthcare-Pharmaceuticals & Medical Research-Pharmaceuticals-Pharmaceuticals
United Kingdom-Industrials-Industrial & Commercial Services-Professional & Commercial Services-Business Support Services
United Kingdom-Technology-Software & IT Services-Software & IT Services-IT Services & Consulting
United Kingdom-Technology-Software & IT Services-Software & IT Services-Online Services
United Kingdom-Technology-Software & IT Services-Software & IT Services-Software
United Kingdom-Technology-Technology Equipment-Semiconductors & Semiconductor Equipment-Semiconductors
Panel B: Control
Canada-Energy-Renewable Energy-Renewable Energy-Renewable Energy Equipment & Services
Canada-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
Canada-Technology-Software & IT Services-Software & IT Services-IT Services & Consulting

Canada-Technology-Software & IT Services-Software & IT Services-Software

Canada-Technology-Technology Equipment-Communications & Networking-Communications & Networking

Canada-Technology-Technology Equipment-Semiconductors & Semiconductor Equipment-Semiconductors

China (Mainland)-Academic & Educational Services-Academic & Educational Services-Miscellaneous Educational Service Providers-Miscellaneous Educational Service Providers

China (Mainland)-Healthcare-Healthcare Services & Equipment-Healthcare Equipment & Supplies-Medical Equipment, Supplies & Distribution

China (Mainland)-Healthcare-Healthcare Services & Equipment-Healthcare Providers & Services-Healthcare Facilities & Services

China (Mainland)-Healthcare-Pharmaceuticals & Medical Research-Pharmaceuticals-Pharmaceuticals

China (Mainland)-Industrials-Industrial & Commercial Services-Professional & Commercial Services-Business Support Services

China (Mainland)-Technology-Software & IT Services-Software & IT Services-IT Services & Consulting

China (Mainland)-Technology-Software & IT Services-Software & IT Services-Online Services

China (Mainland)-Technology-Software & IT Services-Software & IT Services-Software

China (Mainland)-Technology-Technology Equipment-Computers, Phones & Household Electronics-Computer Hardware

India-Financials-Banking & Investment Services-Banking Services-Consumer Lending

India-Healthcare-Healthcare Services & Equipment-Healthcare Providers & Services-Healthcare Facilities & Services

India-Technology-Software & IT Services-Software & IT Services-Online Services

India-Technology-Software & IT Services-Software & IT Services-Software

United States-Academic & Educational Services-Academic & Educational Services-Professional & Business Education-Professional & Business Education

United States-Basic Materials-Chemicals-Chemicals-Commodity Chemicals

United States-Basic Materials-Chemicals-Chemicals-Specialty Chemicals

United States-Consumer Cyclicals-Automobiles & Auto Parts-Automobiles & Auto Parts-Auto & Truck Manufacturers

United States-Consumer Cyclicals-Automobiles & Auto Parts-Automobiles & Auto Parts-Auto, Truck & Motorcycle Parts

United States-Consumer Cyclicals-Cyclical Consumer Products-Leisure Products-Recreational Products

United States-Consumer Cyclicals-Cyclical Consumer Products-Textiles & Apparel-Apparel & Accessories

United States-Consumer Cyclicals-Cyclical Consumer Services-Media & Publishing-Advertising & Marketing

United States-Consumer Cyclicals-Cyclical Consumer Services-Media & Publishing-Entertainment Production

United States-Consumer Cyclicals-Retailers-Diversified Retail-Department Stores

United States-Consumer Cyclicals-Retailers-Specialty Retailers-Miscellaneous Specialty Retailers

United States-Consumer Non-Cyclicals-Food & Beverages-Food & Tobacco-Food Processing

United States-Consumer Non-Cyclicals-Personal & Household Products & Services-Personal & Household Products & Services-Personal Products
United States-Consumer Non-Cyclicals-Personal & Household Products & Services-Personal & Household Products & Services-Personal Services
United States-Energy-Renewable Energy-Renewable Energy-Renewable Energy Equipment & Services
United States-Financials-Banking & Investment Services-Banking Services-Consumer Lending
United States-Financials-Banking & Investment Services-Investment Banking & Investment Services-Investment Management & Fund Operators
United States-Healthcare-Healthcare Services & Equipment-Healthcare Equipment & Supplies-Advanced Medical Equipment & Technology
United States-Healthcare-Healthcare Services & Equipment-Healthcare Equipment & Supplies-Medical Equipment, Supplies & Distribution
United States-Healthcare-Healthcare Services & Equipment-Healthcare Providers & Services-Healthcare Facilities & Services
United States-Healthcare-Pharmaceuticals & Medical Research-Biotechnology & Medical Research-Biotechnology & Medical Research
United States-Healthcare-Pharmaceuticals & Medical Research-Pharmaceuticals-Pharmaceuticals
United States-Industrials-Industrial & Commercial Services-Professional & Commercial Services-Business Support Services
United States-Industrials-Industrial & Commercial Services-Professional & Commercial Services-Environmental Services & Equipment
United States-Industrials-Industrial Goods-Machinery, Tools, Heavy Vehicles, Trains & Ships-Electrical Components & Equipment
United States-Industrials-Industrial Goods-Machinery, Tools, Heavy Vehicles, Trains & Ships-Industrial Machinery & Equipment
United States-Technology-Software & IT Services-Software & IT Services-IT Services & Consulting
United States-Technology-Software & IT Services-Software & IT Services-Online Services
United States-Technology-Software & IT Services-Software & IT Services-Software
United States-Technology-Technology Equipment-Communications & Networking-Communications & Networking
United States-Technology-Technology Equipment-Computers, Phones & Household Electronics-Computer Hardware
United States-Technology-Technology Equipment-Computers, Phones & Household Electronics-Household Electronics
United States-Technology-Technology Equipment-Electronic Equipment & Parts-Electronic Equipment & Parts
United States-Technology-Technology Equipment-Semiconductors & Semiconductor Equipment-Semiconductors
United States-Technology-Telecommunications Services-Telecommunications Services-Integrated Telecommunications Services
United States-Technology-Telecommunications Services-Telecommunications Services-Wireless Telecommunications Services
United States-Utilities-Utilities-Electric Utilities & IPPs-Electric Utilities

Table 5. Results for DID TWFE Regressions. Analysis at Country-Industry Level.

This table reports the DID TWFE analysis examining the impact of the AIFMD on HHI computed at the country–industry level over the period 2008–2018. Two specifications are reported: ln(HHI) and logit(HHI). No control variables are included. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors, reported in parentheses, are clustered at the country–industry level.

	Ln(HHI)	**Logit(HHI)**
Treated x Post	-0.284**	-0.342**
	(0.124)	(0.140)
Num.Obs.	825.00	741.00
R2	0.65	0.64
R2 Adj.	0.61	0.59
R2 Within	0.02	0.02
R2 Within Adj.	0.02	0.02
AIC	1,265.40	1,423.40
BIC	1,670.90	1,819.70
RMSE	0.47	0.56
Std.Errors	by: unit	by: unit
FE: unit	X	X
FE: year	X	X

Table 6. Results for Event Studies (DID TWFE) Regressions. Analysis at Country-Industry Level.

This table reports the event-study analysis examining the impact of the AIFMD on HHI computed at the country–industry level over the period 2008–2018. Two specifications are reported: ln(HHI) and logit(HHI). No control variables are included. The interaction term for 2012 is omitted to avoid multicollinearity and therefore serves as the reference year. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors, reported in parentheses, are clustered at the country–industry level.

	Ln(HHI)	**Logit(HHI)**
Years to Treatment = -5 x Unit Treated	0.345	0.254
	(0.229)	(0.261)
Years to Treatment = -4 x Unit Treated	0.236	0.407
	(0.218)	(0.278)
Years to Treatment = -3 x Unit Treated	-0.025	-0.059
	(0.208)	(0.263)
Years to Treatment = -2 x Unit Treated	-0.085	-0.217
	(0.196)	(0.232)
Years to Treatment = 0 x Unit Treated	0.054	-0.062
	(0.162)	(0.171)
Years to Treatment = 1 x Unit Treated	-0.211	-0.360**
	(0.148)	(0.162)
Years to Treatment = 2 x Unit Treated	0.067	-0.003
	(0.175)	(0.214)
Years to Treatment = 3 x Unit Treated	-0.258*	-0.385*
	(0.153)	(0.212)
Years to Treatment = 4 x Unit Treated	-0.479***	-0.449**

	(0.158)	(0.217)
Years to Treatment = 5 x Unit Treated	-0.310*	-0.323
	(0.18)	(0.221)
Num.Obs.	825	741
R2	0.663	0.645
R2 Adj.	0.62	0.594
R2 Within	0.045	0.04
R2 Within Adj.	0.032	0.025
AIC	1259.4	1424.1
BIC	1707.4	1861.8
RMSE	0.46	0.56
Std.Errors	by: unit	by: unit
FE: unit	X	X
FE: year	X	X

Figure 1. Event Study Plot for Ln(HHI).

This figure presents the event-study plot (DID TWFE) for ln(HHI). The reference year (2012) is indicated by the dashed vertical line. Error bands represent the 95% confidence interval.

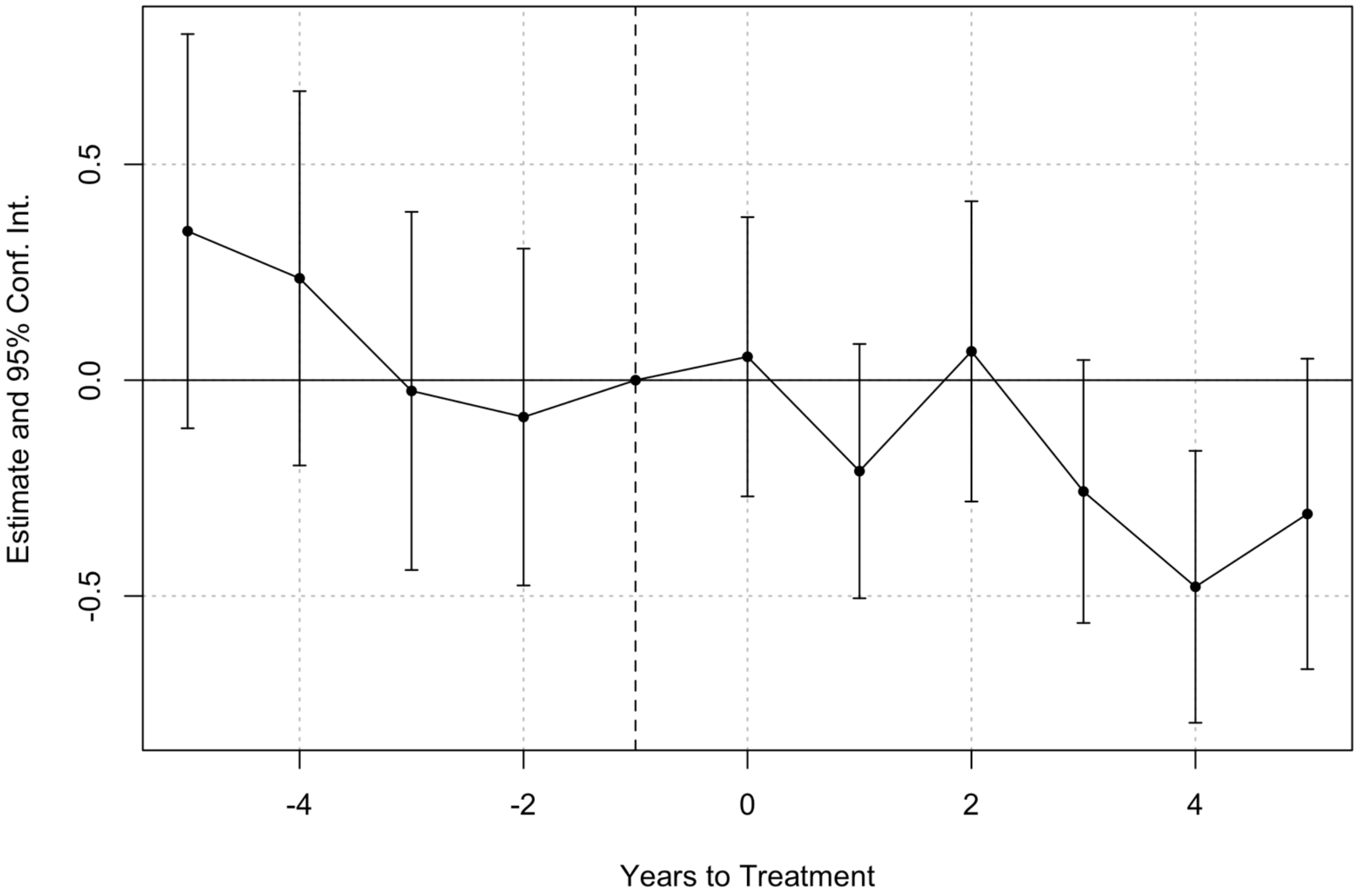


Figure 2. Event Study Plot for Logit(HHI).

This figure presents the event-study plot (DID TWFE) for Logit(HHI). The reference year (2012) is indicated by the dashed vertical line. Error bands represent the 95% confidence interval.

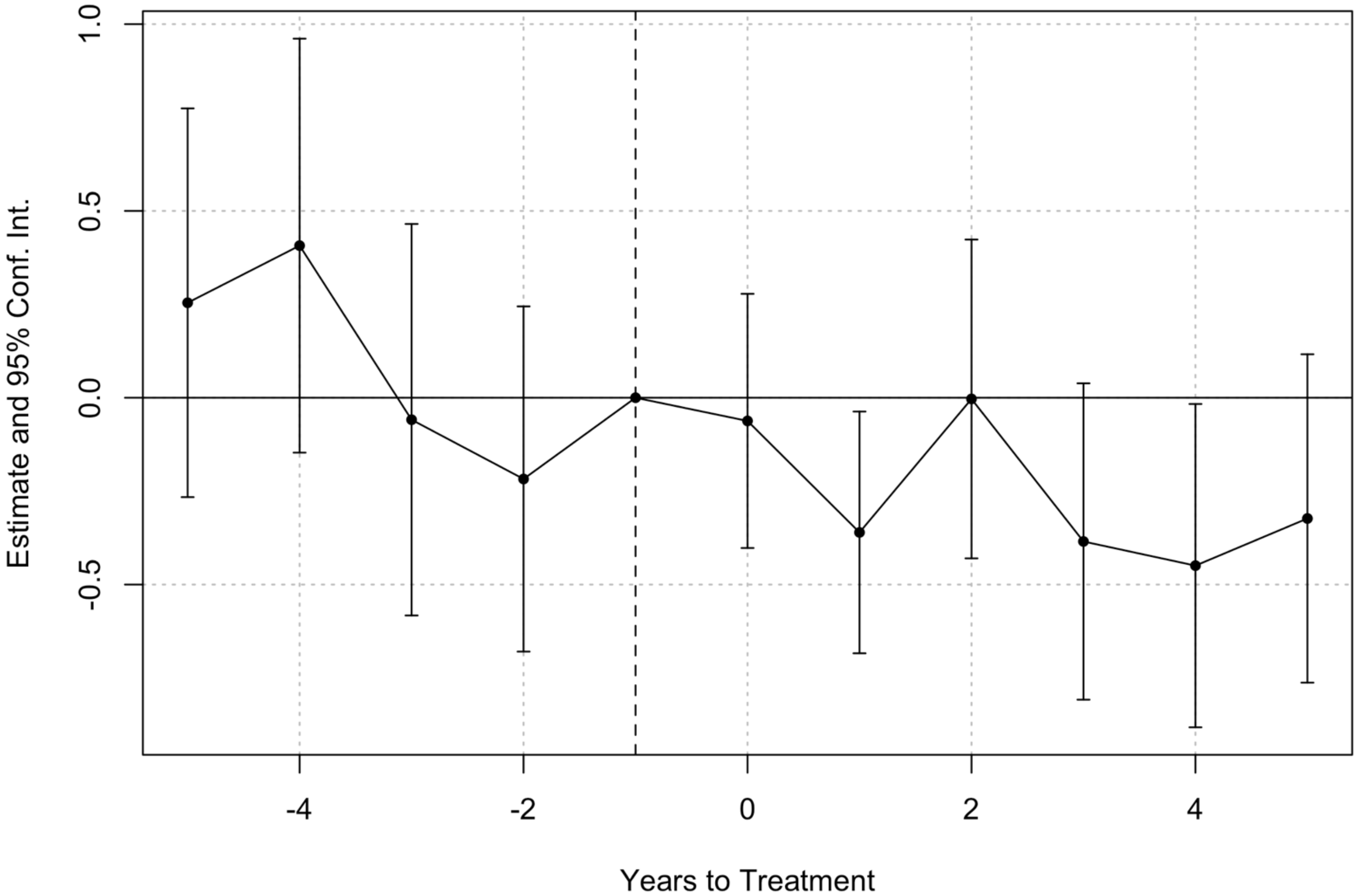


Table 7. Results for SDID Regressions. Analysis at Country-Industry Level.

This table reports the SDID analysis examining the impact of the AIFMD on HHI computed at the country–industry level over the period 2008–2018. Two specifications are reported: ln(HHI) and logit(HHI). No control variables are included. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors are reported in parentheses.

	Ln(HHI)	Logit(HHI)
Treated x Post	-0.26812**	-0.53943***
	(0.11795)	(0.18115)
t-Statistic	-2.273	-2.978
95% CI Lower	-0.49931	-0.89449
95% CI Upper	-0.03693	-0.18438
N Treated Units	19	10
N Control Units	75	38
N Total	94	48

Table 8. Robustness Checks for the Analysis at Country-Industry Level.

This table presents the results of several robustness checks conducted for our country–industry level analysis. Panel A reports results using alternative start and end years, corresponding to different time windows. Panel B reports leave-one-out (LOO) tests for the treated group, in which one country is excluded from the analysis in each specification. Panel C reports results from the stepwise addition of countries to the treated group while maintaining the initial cohort. Panel D reports LOO tests for the control group. Panel E reports results from the stepwise addition of countries to the control group. No control variables are included. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors, reported in parentheses, are clustered at the country–industry level.

Specification	Outcome	Coefficient	Std. Error	t-Statistic	p-Value	Significance	N
Panel A: Different Time Windows							
Start Year = 2007	Ln(HHI)	-0.36295	0.11744	-3.091	0.0021	***	876
Start Year = 2007	Logit(HHI)	-0.40286	0.14049	-2.867	0.0042	***	789
Start Year = 2006	Ln(HHI)	-0.39122	0.12536	-3.121	0.0019	***	884
Start Year = 2006	Logit(HHI)	-0.44029	0.14823	-2.97	0.0031	***	809
Start Year = 2005	Ln(HHI)	-0.34982	0.13504	-2.59	0.0098	***	826
Start Year = 2005	Logit(HHI)	-0.41552	0.16662	-2.494	0.0128	**	758
End Year = 2019	Ln(HHI)	-0.28953	0.12501	-2.316	0.0208	**	864
End Year = 2019	Logit(HHI)	-0.3549	0.14278	-2.486	0.0131	**	786
Panel B: Leave-One-Out (LOO) Tests for the Treated Group							
LOO Treated: United Kingdom	Ln(HHI)	-0.17871	0.13669	-1.307	0.1915		726
LOO Treated: United Kingdom	Logit(HHI)	-0.23289	0.16381	-1.422	0.1556		650
LOO Treated: France	Ln(HHI)	-0.34565	0.14389	-2.402	0.0165	**	770
LOO Treated: France	Logit(HHI)	-0.38856	0.15692	-2.476	0.0135	**	687
LOO Treated: Germany	Ln(HHI)	-0.29921	0.13295	-2.251	0.0247	**	792
LOO Treated: Germany	Logit(HHI)	-0.34131	0.14866	-2.296	0.022	**	709
LOO Treated: Sweden	Ln(HHI)	-0.26699	0.12821	-2.082	0.0376	**	814
LOO Treated: Sweden	Logit(HHI)	-0.33363	0.14374	-2.321	0.0206	**	731

LOO Treated: Switzerland	Ln(HHI)	-0.2961	0.1287	-2.301	0.0217	**	814
LOO Treated: Switzerland	Logit(HHI)	-0.37578	0.13978	-2.688	0.0073	***	731
Panel C: Stepwise Addition of Countries to the Treated Group							
Add Treated: Netherlands	Ln(HHI)	-0.28354	0.12418	-2.283	0.0227	**	825
Add Treated: Netherlands	Logit(HHI)	-0.34198	0.13951	-2.451	0.0145	**	741
Add Treated: Denmark	Ln(HHI)	-0.28354	0.12418	-2.283	0.0227	**	825
Add Treated: Denmark	Logit(HHI)	-0.34198	0.13951	-2.451	0.0145	**	741
Add Treated: Ireland	Ln(HHI)	-0.28354	0.12418	-2.283	0.0227	**	825
Add Treated: Ireland	Logit(HHI)	-0.34198	0.13951	-2.451	0.0145	**	741
Panel D: Leave-One-Out (LOO) Tests for the Control Group							
LOO Control: United States	Ln(HHI)	-0.28266	0.15192	-1.861	0.0635	*	418
LOO Control: United States	Logit(HHI)	-0.37044	0.21203	-1.747	0.0815	*	356
LOO Control: China (Mainland)	Ln(HHI)	-0.24446	0.1288	-1.898	0.0581	*	726
LOO Control: China (Mainland)	Logit(HHI)	-0.25498	0.14195	-1.796	0.0729	*	663
LOO Control: Canada	Ln(HHI)	-0.29958	0.12501	-2.397	0.0168	**	759
LOO Control: Canada	Logit(HHI)	-0.37595	0.14171	-2.653	0.0082	***	693
LOO Control: India	Ln(HHI)	-0.30376	0.12569	-2.417	0.0159	**	781
LOO Control: India	Logit(HHI)	-0.37967	0.14074	-2.698	0.0071	***	708
LOO Control: Japan	Ln(HHI)	-0.28354	0.12418	-2.283	0.0227	**	825
LOO Control: Japan	Logit(HHI)	-0.34198	0.13951	-2.451	0.0145	**	741
Panel E: Stepwise Addition of Countries to the Control Group							
Add Control: Australia	Ln(HHI)	-0.26562	0.12491	-2.127	0.0337	**	836
Add Control: Australia	Logit(HHI)	-0.3251	0.13956	-2.329	0.0201	**	751
Add Control: Singapore	Ln(HHI)	-0.28354	0.12418	-2.283	0.0227	**	825
Add Control: Singapore	Logit(HHI)	-0.34198	0.13951	-2.451	0.0145	**	741

Add Control: Israel	Ln(HHI)	-0.27396	0.1245	-2.2	0.028	**	847
Add Control: Israel	Logit(HHI)	-0.31511	0.13974	-2.255	0.0244	**	760

Table 9. Results for DID TWFE Regressions. Analysis at the Startup-Level.

This table presents the results for our startup-level level analysis. No control variables are included. Our sample includes Seed and Early Stage Startups. These were matched using Propensity Score Matching. Matching is performed on average equity funding ("ind_avg_equity_pre") and average investor per deal ("ind_avg_investors_pre"). These are calculated at the industry level and pre-treatment (thus, they are time-invariant). In addition, we match on startup age at funding round ("age_at_funding"). ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors, reported in parentheses, are clustered at the country–industry level.

	Ln(Round_Equity)	**Ln(No_Investors)**	**Ln(Avg_Investment)**
Treated x Post	-0.323**	-0.069**	-0.251**
	(0.13)	(0.03)	(0.13)
Num.Obs.	2,725.00	2,725.00	2,725.00
R2	0.16	0.13	0.14
R2 Adj.	0.09	0.06	0.06
R2 Within	0.00	0.00	0.00
R2 Within Adj.	0.00	0.00	0.00
AIC	9,769.40	3,594.10	9,550.00
BIC	11,069.70	4,894.40	10,850.30
RMSE	1.34	0.43	1.29
Std.Errors	by: country_industry	by: country_industry	by: country_industry
FE: country^industry	X	X	X
FE: year	X	X	X

Figure 3. Event Study Plot for Ln(Round_Equity).

This figure presents the event-study plot (DID TWFE) for Ln(Round_Equity). The reference year (2012) is indicated by the dashed vertical line. Error bands represent the 95% confidence interval.

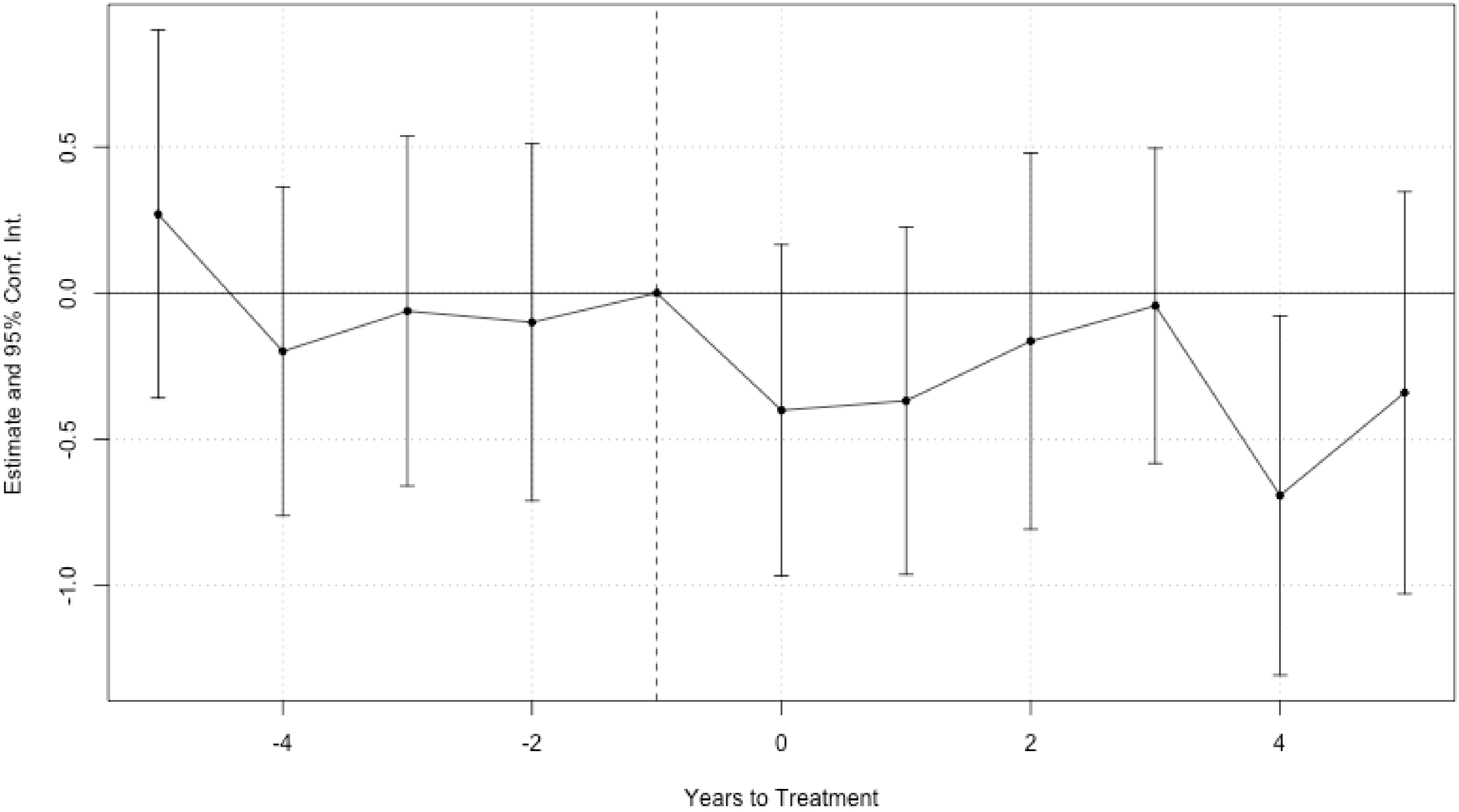


Figure 4. Event Study Plot for Ln(No_Investors).

This figure presents the event-study plot (DID TWFE) for Ln(No_Investors). The reference year (2012) is indicated by the dashed vertical line. Error bands represent the 95% confidence interval.

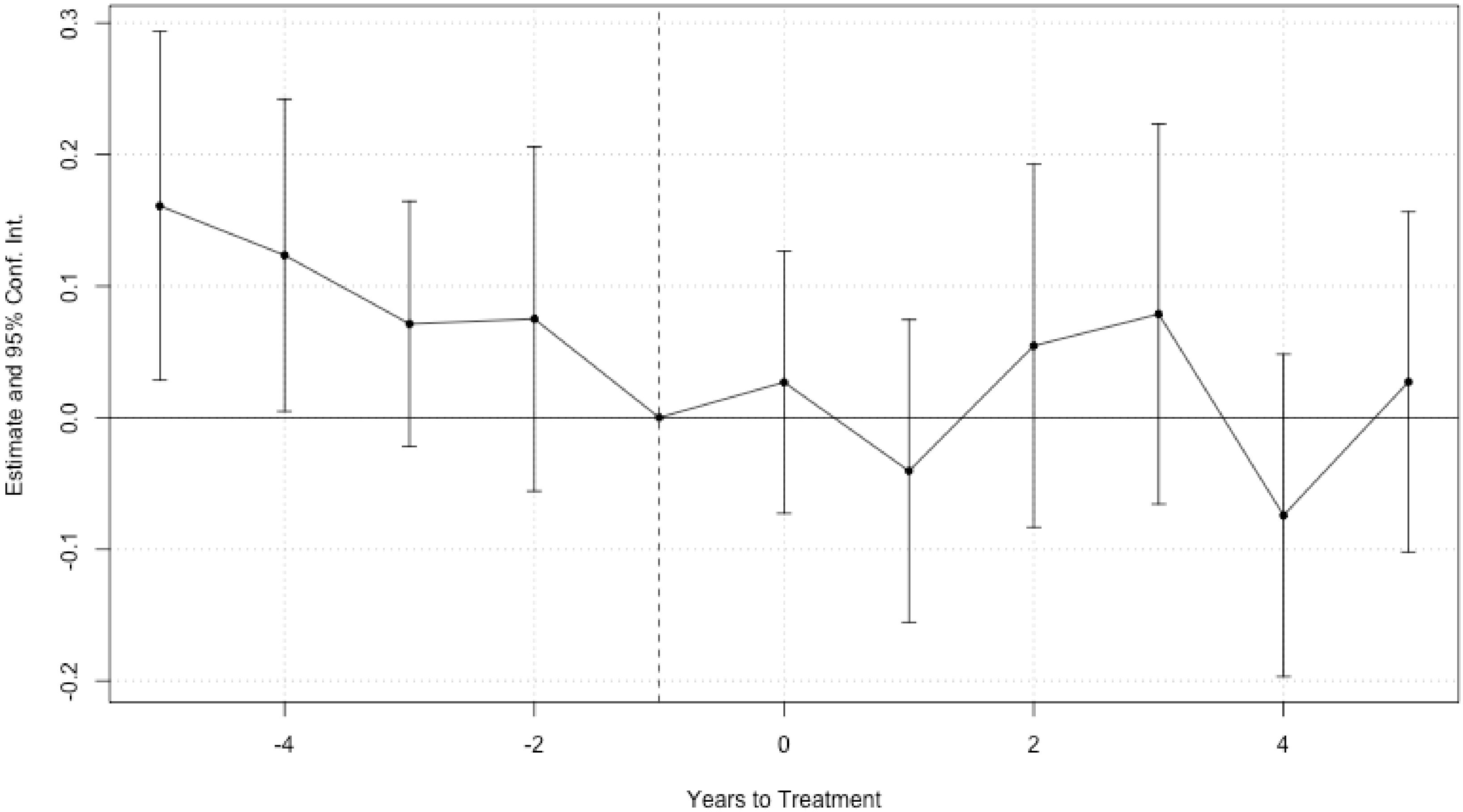


Figure 5. Event Study Plot for Ln(Avg_Investment).

This figure presents the event-study plot (DID TWFE) for Ln(Avg_Investment). The reference year (2012) is indicated by the dashed vertical line. Error bands represent the 95% confidence interval.

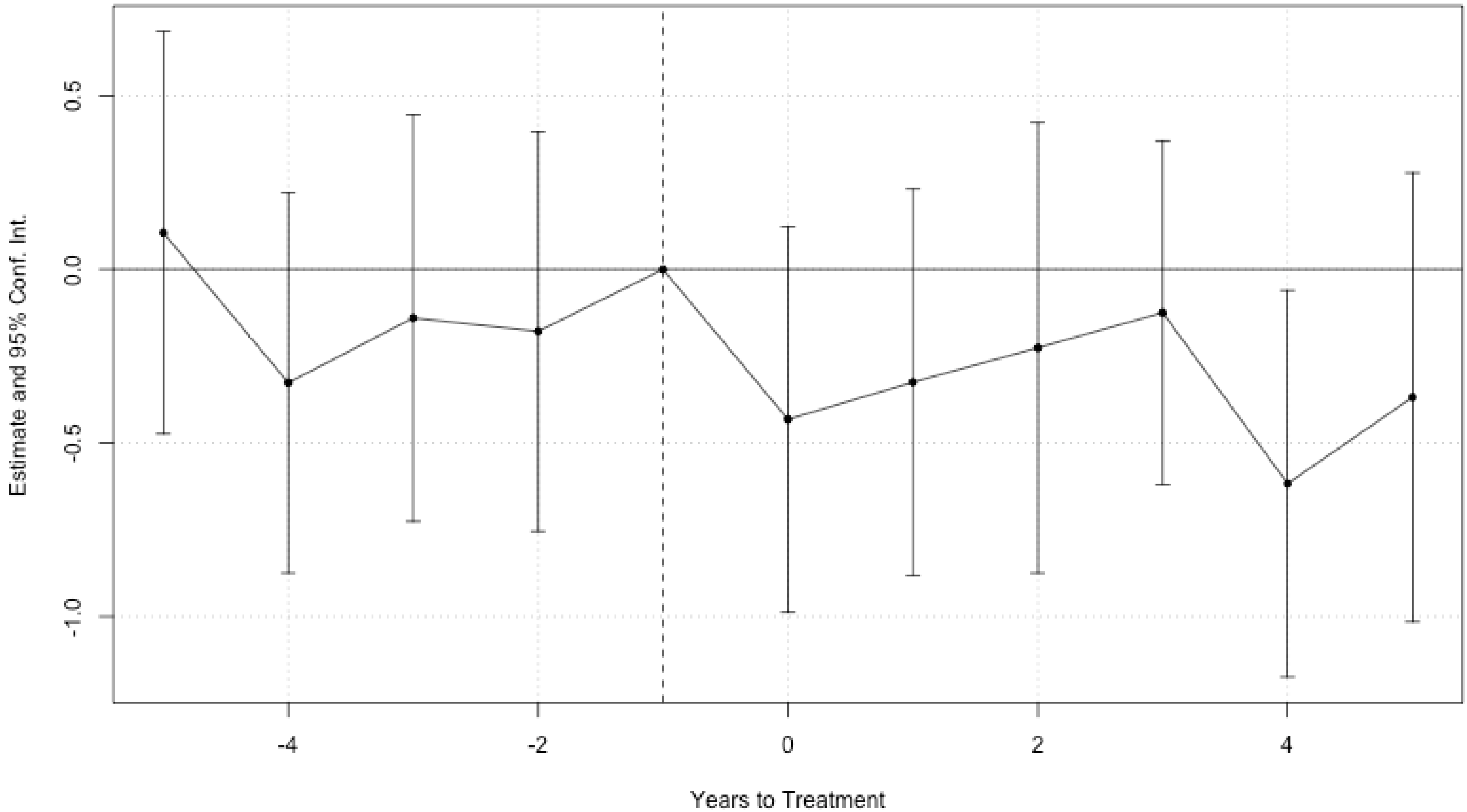


Table 10. Robustness Checks for the Analysis at Startup Level.

This table presents the results of several robustness checks conducted for the startup-level analysis. Panel A reports results using alternative matching strategies, namely Unmatched, Coarsened Exact Matching (CEM), and Entropy Balancing (Entropy Bal.). Panel B reports results from an alternative specification that matches startups using only the average amount of equity raised pre-treatment, calculated at the industry level ("ind_avg_equity_pre"), and the startup's age at funding ("age_at_funding"). Panel C reports results from an alternative specification that matches startups using only the average number of investors per round pre-treatment, calculated at the industry level ("ind_avg_investors_pre"), and the startup's age at funding ("age_at_funding"). No control variables are included. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. Standard errors are clustered at the country–industry level.

Sample	Outcome	Coefficient	Std. Error	P-Value	Significance	N
Panel A: Different Matching Strategies						
Unmatched	ln(Round Equity)	-0.2406	0.116	0.0392	**	6,966
Unmatched	ln(No. Investors)	-0.0899	0.0231	0.0001	***	6,966
Unmatched	ln(Avg Investment)	-0.1476	0.1164	0.2061		6,966
CEM	ln(Round Equity)	-0.233	0.1188	0.0511	*	6,919
CEM	ln(No. Investors)	-0.0931	0.0239	0.0001	***	6,919
CEM	ln(Avg Investment)	-0.137	0.1193	0.2522		6,919
Entropy Bal.	ln(Round Equity)	-0.2263	0.1153	0.051	*	6,966
Entropy Bal.	ln(No. Investors)	-0.093	0.0235	0.0001	***	6,966
Entropy Bal.	ln(Avg Investment)	-0.1301	0.1154	0.2607		6,966
Panel B: Alternative Specification - "ind_avg_equity_pre" + "age_at_funding".						
Unmatched	ln(Round Equity)	-0.2406	0.116	0.0392	**	6,966
Unmatched	ln(No. Investors)	-0.0899	0.0231	0.0001	***	6,966
Unmatched	ln(Avg Investment)	-0.1476	0.1164	0.2061		6,966

PSM	ln(Round Equity)	-0.3262	0.1249	0.0097	***	2,722
PSM	ln(No. Investors)	-0.081	0.0314	0.0106	**	2,722
PSM	ln(Avg Investment)	-0.2427	0.1234	0.0506	*	2,722
CEM	ln(Round Equity)	-0.2307	0.1216	0.059	*	6,949
CEM	ln(No. Investors)	-0.0851	0.0238	0.0004	***	6,949
CEM	ln(Avg Investment)	-0.1425	0.1232	0.2486		6,949
Entropy Bal.	ln(Round Equity)	-0.2282	0.1169	0.0522	*	6,966
Entropy Bal.	ln(No. Investors)	-0.0902	0.0236	0.0002	***	6,966
Entropy Bal.	ln(Avg Investment)	-0.1348	0.1175	0.2528		6,966

Panel C: Alternative Specification - “ind_avg_investors_pre” + “age_at_funding”.

Unmatched	ln(Round Equity)	-0.2406	0.116	0.0392	**	6,966
Unmatched	ln(No. Investors)	-0.0899	0.0231	0.0001	***	6,966
Unmatched	ln(Avg Investment)	-0.1476	0.1164	0.2061		6,966
PSM	ln(Round Equity)	-0.273	0.132	0.0398	**	2,726
PSM	ln(No. Investors)	-0.0631	0.0322	0.0516	*	2,726
PSM	ln(Avg Investment)	-0.209	0.1287	0.106		2,726
CEM	ln(Round Equity)	-0.238	0.1209	0.0501	*	6,941
CEM	ln(No. Investors)	-0.0872	0.0237	0.0003	***	6,941
CEM	ln(Avg Investment)	-0.1481	0.1209	0.2219		6,941
Entropy Bal.	ln(Round Equity)	-0.2277	0.1138	0.0465	**	6,966
Entropy Bal.	ln(No. Investors)	-0.0929	0.0232	0.0001	***	6,966

Entropy Bal.	ln(Avg Investment)	-0.1317	0.1136	0.2474		6,966

The Term Structure of Carbon Return

Ke Wang*

Supervisors: Kevin Aretz† Eirini Konstantinidi‡ Hening Liu§

April 2026

Abstract

We study the term structure of carbon return. Existing evidence on carbon transition risk is almost entirely stock-level and therefore aggregates cash flows across horizons. Using single-stock dividend futures, we show that carbon return is strongly horizon-dependent. In our sample, more carbon-intensive firms earn lower stock realized returns, but this negative carbon return is concentrated in short-maturity dividend strips and weakens rapidly with maturity. At the far end of the term structure, the sign reverses: the residual long-term asset exhibits a positive carbon return in expected returns. These results imply that stock-level carbon return is an aggregation of negative short-end and positive long-end premia, and that carbon risk is not priced as a single maturity-invariant object, but instead displays a pronounced term structure.

Keywords: carbon return, climate finance, dividend strips, term structure, long-term asset

JEL codes: G12, G14, Q54

* The University of Manchester; E-mail: ke.wang-9@postgrad.manchester.ac.uk

† The University of Manchester; E-mail: kevin.aretz@manchester.ac.uk

‡ The University of Manchester; E-mail: eirini.konstantinidi@manchester.ac.uk

§ The University of Manchester; E-mail: Hening.Liu@manchester.ac.uk

I. Introduction

The pricing of carbon transition risk has become a central question in climate finance. A growing literature studies whether brown firms earn higher returns than green firms and whether investors require compensation for exposure to carbon emissions. Yet the evidence remains sharply divided. Bolton and Kacperczyk (2021, 2023) document a positive association between returns and carbon emissions, consistent with a carbon premium. Other studies instead find that green assets outperform, consistent with investor demand for climate hedges, changing tastes, or climate-concern shocks. More recent evidence in Zhang (2025) shows that once emissions are measured in real time and data-release lags are properly accounted for, carbon returns turn negative in the United States and are weak or insignificant globally, especially in developed markets. The literature has therefore moved from asking whether carbon risk is priced at all to asking what exactly stock-level carbon returns are measuring.

A key difficulty is that the existing literature studies the stock as a whole. But a stock is a claim on a sequence of future cash flows. In the absence of arbitrage, its expected return is a value-weighted average of the expected returns on the underlying cash flows. If carbon risk is priced differently across horizons, stock-level returns may mask substantial heterogeneity across the cash-flow term structure. This observation connects the climate-finance literature to an older and broader literature on the timing and pricing of cash flows in asset prices. Lucas (1978) and Gordon (1962) provide the basic present-value logic, while Mehra and Prescott (1985) and Hansen and Singleton (1982, 1983) frame the equity-premium puzzle in terms of aggregate stock returns. van Binsbergen, Brandt, and Koijen (2012) shift attention from the stock market as a whole to the individual terms in the dividend sum and show that short-term dividend strips have higher risk premia than long-term strips. Their paper is the natural modern starting point for the equity term-structure literature.

That literature has since developed along two related directions. One direction studies the aggregate equity term structure. Following van Binsbergen, Brandt, and Koijen (2012), van Binsbergen et al. (2013) and van Binsbergen and Koijen (2017) document further facts on dividend-strip pricing and the term structure of equity returns. More recent work, including Miller (2020), Chen (2020), Giglio, Kelly, and Kozak (2020), Andrews and Gonçalves (2020), Cejnek and Randl (2020), and Gormsen (2021), studies the slope of the equity term structure using alternative methods and data. A parallel direction studies duration in the cross-section of stock returns. Dechow, Sloan, and Soliman (2004) introduce implied equity duration; Da (2009) and Weber (2018) link duration to cross-sectional return patterns; Lettau and Wachter (2007), Croce, Lettau, and Ludvigson (2009), Chen and Li (2018), and Gonçalves (2021) argue that duration helps explain value, profitability, and investment premia. Gormsen and Lazarus (2023) bring these strands together. Using single-stock dividend futures, they show that risk-adjusted returns decline with maturity within firms and, crucially, provide direct identification by holding cash-flow maturity fixed while varying firm characteristics, or vice versa.

The climate-finance literature, by contrast, has largely remained at the stock level. Bolton and Kacperczyk (2021) frame the question through several competing hypotheses: carbon emissions may proxy for priced transition risk, underpricing of carbon risk, or exclusionary preferences resembling sin-stock pricing. Their evidence is strongest for total emissions and emissions changes rather than emissions intensity. Subsequent work expands the debate in several directions. Pastor, Stambaugh, and Taylor (2022), Pedersen, Fitzgibbons, and Pomorski (2021), Ardia et al. (2023), and Alekseev et al. (2022) emphasize green demand, sustainable investing, and climate-concern channels. Dyck et al. (2019) and Gibson Brandon et al. (2022) study responsible institutional investing, while Krueger, Sautner, and Starks (2020) provide survey evidence that climate risk may not be fully priced. Choi, Gao, and Jiang (2020) study short-run price effects when investors revise beliefs about climate change. Görgen et al. (2020), Aswani, Raghunandan, and Rajgopal (2024), and Lindsey, Pruitt, and Schiller (2021) report mixed evidence on carbon returns. Zhang (2025) shows that information timing is central and argues that the previously documented positive carbon premium largely reflects forward-looking firm performance embedded in emissions rather than a true premium in ex ante expected returns.

What is still missing is a direct study of carbon return across cash-flow maturities. Existing carbon-return papers treat the stock as a single object. Existing term-structure papers study horizon-specific pricing, but not carbon risk. This leaves a clear gap. If carbon risk is priced differently across near-term and distant cash flows, then stock-level carbon return is only an aggregation outcome and cannot reveal where along the cash-flow horizon the pricing actually occurs. To our knowledge, no paper has directly measured the term structure of carbon return using traded claims on firm-level cash flows. Gormsen and Lazarus (2023) make precisely this type of identification possible by showing how single-stock dividend futures can be used to hold maturity fixed while varying firm characteristics. We use the same market, but ask a different question: whether carbon exposure itself is priced differently across maturities.

Our paper is also closely related to He, Li, and Zhang (2025), the paper closest in spirit to ours. They argue that equity duration helps reconcile conflicting carbon-return results and document that the sign of the carbon premium depends on whether equity duration is long or short. We view that paper as complementary, but methodologically different. Their design backs out duration from stock-level valuation relations and therefore relies on assumptions about discounting beyond observed horizons. By contrast, our front-end evidence comes directly from traded prices of maturity-specific cash flows. This distinction matters. A quantity derived under an assumed discount-rate structure is not the same object as directly observed horizon-specific cash-flow prices. Our approach is therefore closer to a direct pricing test of the term structure of carbon return, especially at the short end where SSDF prices are observed rather than inferred.

Our main findings are threefold. First, at the stock level, we find that more carbon-intensive firms earn significantly lower future returns in the SSDF-linked sample, in both value-weighted and equal-weighted portfolios, and for both Scope 1 and Scope 2 intensity. This result is quantitatively close to the developed-market evidence in Zhang (2025). Second, at the cash-flow level, the negative carbon return is concentrated in short-maturity dividend strips. At the one- and two-year horizons, the high-minus-low expected return spread is economically large and significantly negative; as maturity increases, both the magnitude and statistical significance weaken, and in some specifications the spread approaches zero or turns positive. Third, at the far end of the term structure, the sign reverses: the residual long-term asset exhibits a positive carbon return in expected returns. The stock-level carbon return is therefore an aggregation of sharply negative short-end premia and positive long-end premia. Carbon return is not a single maturity-invariant object, but a horizon-dependent pricing pattern across cash flows.

This paper contributes to the literature in three ways. First, it contributes to the equity term-structure literature by showing that climate-related pricing is strongly horizon-dependent. Second, it contributes to the climate-finance literature by showing that stock-level carbon return is too coarse an object to reveal the underlying pricing pattern. Third, it contributes methodologically by providing, to our knowledge, the first direct SSDF-based evidence on how carbon return varies across cash-flow maturities. In addition, unlike the standard characteristics studied by Gormsen and Lazarus (2023), carbon exposure continues to generate cross-sectional differences in the pricing of dividend strips even when maturity is held fixed, especially at short horizons.

The remainder of the paper proceeds as follows. Section II describes the data, the matching procedure, and the construction of stock, dividend-strip, and long-term-asset returns. Section III documents the stock-level carbon return in the SSDF-linked sample. Section IV studies short-horizon dividend strips and shows that the negative carbon return is concentrated at the short end of the term structure. Section V turns to the residual long-term asset and shows that the sign reverses at the far end. Section VI shows that the main results are robust to sorting on emissions levels rather than emissions intensity. Section VII concludes.

II. Data and Methodology

This section lays out the empirical framework used to study the term structure of carbon return. It first introduces the cash-flow decomposition and defines the stock, dividend-strip, and long-term-asset objects used in the analysis. It then describes the data sources, the cross-database matching procedure, and the institutional features of the single-stock dividend futures market. Finally, it explains the construction of strip expected returns, CAPM alphas, and the residual long-term asset, and presents the final sample coverage and summary statistics.

A. Background

To study the term structure of carbon return, the relevant object is not only the stock as a whole, but also the sequence of cash flows underlying the stock. If carbon risk is priced differently across horizons, aggregate stock returns may mask substantial heterogeneity across cash-flow maturities. This paper therefore studies three related objects: the stock itself, short-horizon dividend strips, and the residual long-term asset after stripping out near-term cash flows.

In the absence of arbitrage, the value of a stock can be written as the present value of its future dividends:

$$P_{i,t} = \sum_{h=1}^{\infty} E_t\left[M_{t\to t+h}\, D_{i,t+h}\right] \equiv \sum_{h=1}^{\infty} P_{i,d,t}^{(h)} \tag{1}$$

where $P_{i,t}$ is the stock price of firm i at time t, $D_{i,t+h}$ is the dividend paid at horizon $t+h$, $M_{t\to t+h}$ is the stochastic discount factor, and $P_{i,d,t}^{(h)}$ denotes the price of a dividend claim with maturity of h years. Thus, a stock can be viewed as a portfolio of dividend strips with different maturities.

This decomposition implies that the expected return on a stock can be written as the value-weighted expected return on all of its future cash flows:

$$E_t\left[r_{i,t+1}\right] = \sum_{h=1}^{\infty} w_{i,t}^{h}\, E_t\left[r_{i,t+1}^{h}\right] \tag{2}$$

where $r_{i,t+1}^{h}$ is the one-period excess return on the $t+h$ cash flow, and $w_{i,t}^{h}$ is its ex ante relative present value in the stock. The same linear aggregation applies not only to expected returns, but also to CAPM betas, CAPM alphas, and exposures or alphas under any linear factor model. Therefore, if carbon return varies across maturities, the stock-level carbon return is not a single maturity-invariant object, but an aggregation of carbon premia at different horizons.

Empirically, dividend strips make this decomposition feasible. One approach is to back out dividend strip prices through no-arbitrage restrictions on options, as in van Binsbergen, Brandt, and Koijen (2012). However, this approach has been argued to be potentially affected by microstructure noise, taxes, dealer funding costs, and liquidity frictions (Boguth et al., 2012; Schulz, 2016; Song, 2016; Bansal et al., 2019). For this reason, this paper relies on dividend futures instead. In the aggregate market setting, prior studies use index dividend futures to study the term structure of equity cash flows (van Binsbergen et al., 2013; van Binsbergen and Koijen, 2017). Single-stock dividend futures(SSDF) are the firm-level analogue of these contracts, and Gormsen and Lazarus (2023) show how they can be used to study horizon-specific pricing at the individual-firm level.

A single-stock dividend future is a standardized contract written on the ordinary cash dividends paid by an individual firm over a given calendar year. More details of these contracts are discussed in Section II.C.

Let $f_{i,d,t}^{(h)}$ denote the futures price at time t on firm i 's dividend paid in year $t+h$. Under no arbitrage, the corresponding spot price of the dividend strip is

$$P_{i,d,t}^{(h)} = f_{i,d,t}^{(h)} e^{-h y_t^{(h)}} \tag{3}$$

where $y_t^{(h)}$ is the log-yield on an h -year default-free zero-coupon bond. Single-stock dividend futures therefore provide direct market prices of horizon-specific dividend claims.

In practice, observable strip maturities for individual stocks are short, typically extending up to five years. Dividend futures therefore identify mainly the front end of the term structure. To shed light on the far end, this paper further constructs a residual long-term asset by stripping near-term dividend claims out of the stock. Further details on the construction of the long-term asset are provided in Section II.D below. Taken together, the stock, short-horizon dividend strips, and the long-term asset form the three empirical objects used to trace how carbon return is priced across cash-flow horizons.

B. Data Sources and Information Integration

B.1. Single-Stock Dividend Futures

We obtain historical SSDF header information from Deutsche Börse, the owner of the Eurex Exchange on which these contracts trade. The header files contain daily information on all product IDs that have ever been traded in the SSDF market. We then use this comprehensive product list to retrieve daily contract-level data from Bloomberg.[1]

In the raw data, each observation corresponds to a contract on a given date and includes the settlement price, trading volume, open interest, contract size, contract currency, contract

[1] *By contrast, the product list downloaded directly from the Eurex website contains only the product IDs active on the download date and may omit contracts that traded historically but were later discontinued. Using the Deutsche Börse header files therefore helps avoid survivorship bias in the SSDF sample.*

maturity date, and the International Securities Identification Number (ISIN) of the underlying firm. Before cleaning, the raw sample spans 2010 to 2025 and covers 323 underlying firms. The SSDF universe is concentrated in large developed-market firms, mainly in Europe, and spans seven contract currencies: EUR, USD, GBP, DKK, SEK, CHF, and NOK. We discuss the nature of the SSDF market, and the data handling procedures in detail in Section II.C and the Appendix.

B.2. Analyst Dividend Expectations

We match SSDF prices to expected dividends from the I/B/E/S Summary History files, combining the U.S. and international I/B/E/S universes. At each month-end, we use the point-in-time consensus forecast of annual dividends per share for a given firm and calendar year, measured by the median analyst forecast.[2]

B.3. Carbon Emissions

Firm-level carbon emissions data come from S&P Trucost, which reports annual greenhouse gas emissions in tons of carbon dioxide equivalent (tCO2e). Our carbon data cover the 2005 to 2020 financial years in the global sample.[3]

The analysis focuses on Scope 1 and Scope 2 emissions. Scope 1 emissions are direct emissions from sources owned or controlled by the firm, whereas Scope 2 emissions are indirect emissions from purchased electricity, steam, heating, and cooling. Emissions intensity, defined as emissions scaled by sales, is the baseline sorting variable, because carbon emissions scale with firms' operations and intensity is therefore more informative for cross-firm comparisons. To benchmark against the literature, the paper also considers log emissions levels and emissions growth in additional tests.

B.4. Stock Prices and Firm Fundamentals

Monthly stock prices and annual accounting variables are obtained from Compustat monthly and annual files, using both the Global and North America databases to cover the full firm universe. These data are used to construct stock return series in the baseline tests and to describe the sample in the summary statistics and representativeness analysis.

B.5. Data Matching and Information Timing

Following Gormsen and Lazarus (2023), we first link the SSDF universe to Compustat by matching the ISIN of the firm underlying each SSDF contract to Compustat GVKEY. Before matching to analyst expectations, we exclude SSDF observations with zero open interest. An open interest of zero means that there is no unsettled contract outstanding in the market, that is, no investor currently holds the claim. This filter is particularly relevant for long-maturity dividend futures early in their life cycle, when listed contracts may exist but have not yet become active. The remaining SSDF observations are then matched to I/B/E/S dividend expectations at monthly frequency for the same firm and claim year. In this step, 49,498 out of 64,019 SSDF observations are successfully matched to analyst expectations. The resulting matched panel is used to construct expected returns on short-horizon dividend strips: the SSDF price identifies the market value of the claim, and the I/B/E/S forecast provides the corresponding expected dividend level.

---

[2] *Analyst expectations may be subject to systematic bias. The median is therefore preferred to the mean as a more robust measure of consensus forecasts.*

[3] *More recent emissions data exist but are not available to us.*

Annual emissions data and annual accounting data are then merged into the monthly firm panel based on their respective data release dates, following Zhang (2025). Emissions data are known to be subject to substantial reporting delays. In our setting, Trucost emissions observations enter the panel only from the vendor's effective date onward, and Compustat annual variables enter only after the corresponding report date.[4] Each annual observation is then carried forward month by month until a new observation becomes available. This timing convention ensures that carbon emissions and firm fundamentals are known at portfolio formation and helps avoid look-ahead bias in the portfolio analysis.

Because SSDF coverage is limited, the final sample begins in January 2012 to ensure a meaningful cross-sectional size in the dividend-strip portfolio analysis. In addition, because our carbon emissions data end in financial year 2020, we truncate the sample after December 2022 to avoid using excessively stale sorting variables. The final sample covers 189 firms from January 2012 to December 2022, yielding 13,226 firm-month observations. More detail on sample coverage and summary statistics is provided in Section II.E below.

C. The Single-Stock Dividend Futures Market

Throughout the paper, we focus on annual single-stock dividend futures contracts.[5] These contracts are claims on the ordinary cash dividends paid by an individual firm over a given calendar year, without reinvestment.[6] For example, the 2025 dividend future on Adidas gives the buyer the right to the dividends paid by Adidas during the 2025 calendar year. SSDF prices therefore reflect the market value of firm-level cash flows paid within a given year, which makes them well suited to studying the prices and returns of cash flows at different horizons for individual firms.

Single-stock dividend futures first traded as dividend swaps in the over-the-counter market in the early 2000s (Manley and Mueller-Glissmann, 2008). Since 2010, they have traded as standardized products on the Eurex Exchange. The market has expanded substantially over time. It began with 50 contracts on Euro Stoxx 50 constituent stocks, and by the end of 2025 covered roughly 300 firms. Figure 1 plots the evolution of the SSDF market. The left axis shows the number of underlying firms, while the right axis shows the average number of December contracts per firm. Shortly after launch, the market expanded to include 10 Swiss SSDFs and 13 UK SSDFs. Between March 2014 and July 2019, there were three major expansions, and after 2020 the market continued to expand through smaller but more frequent additions. At the same time, the observable maturity structure remains short. For most firms, only five successive annual contracts are available at a point in time.[7]

Insert Figure 1 here

Although SSDFs are exchange-traded products, the market retains some features of over-the-counter trading. In particular, a substantial share of activity is executed through OTC channels and subsequently cleared through Eurex facilities for risk-management purposes (Gormsen and Lazarus, 2023). As a result, the settlement prices we observe are end-of-day prices against which positions are cleared in the risk-management system. These prices may reflect traded prices, but in some cases may also rely on a combination of quotes and

---

[4] *Trucost reviewed and updated all pre-2008 observations in May 2009. We therefore exclude all emissions data prior to 2008 because they are backfilled.*

[5] *In unreported results, we continue to carry forward 2020 emissions beyond December 2022. This does not change the main results; these results are available upon request.*

[6] *Since 2020, Eurex has also introduced quarterly contracts for a selected group of firms.*

[7] *More precisely, the contract covers dividends paid from the last settlement date of the previous year to the settlement date of the current year. For annual contracts, the settlement date is the third Friday of December.*

proprietary models. Following Gormsen and Lazarus (2023), we therefore treat trading activity as an important signal of price informativeness and track prices in calendar time, updating them only on days when we observe trading volume in the market, to ensure that the prices used in the analysis are based on traded prices.

Liquidity also varies sharply across contracts, and prices can become stale for less actively traded names. This heterogeneity is evident in our sample. Some contracts trade frequently and accumulate substantial open interest, whereas others trade only sporadically and may go months without a transaction. Such features make SSDFs ill-suited to high-frequency return analysis. In our sample, about 66.7% of monthly realized strip returns are zero. We therefore avoid relying on realized strip returns whenever possible. Instead, the main analysis focuses on expected returns constructed from strip prices and analyst dividend expectations.

D. Identification from Dividend Strips

The empirical challenge in studying the term structure of carbon return is that the stock aggregates cash flows at all horizons. Single-stock dividend futures identify the front end of the cash-flow term structure, but only for a limited set of near-term maturities. We therefore proceed in two steps. First, we use observed strip prices together with analyst dividend expectations to measure expected returns and CAPM alphas on short-horizon dividend strips. Second, because dividend futures are available only for relatively short maturities, we construct a residual long-term asset to capture the far end of the term structure. Taken together, these two objects allow us to trace how carbon return varies across cash-flow horizons.

D.1. Expected Returns and CAPM Alphas on Short-Horizon SSDF

We begin with the short end of the term structure. Following Gormsen and Lazarus (2023), we combine prices of single-stock dividend futures with analysts' expectations of future dividends to measure expected returns on dividend strips. Specifically, for firm i, let $f_{i,d,t}^{(h)}$ denote the time-t futures price of the dividend paid at horizon $t+h$, and let $\mathbb{E}_t[D_{i,t+h}]$ denote the corresponding point-in-time expectation of that dividend. We measure the expected return on the h -year dividend strip as the expected yield-to-maturity,

$$\mathbb{E}_t\left[r_{i,t+h}^h\right]=\left(\frac{\mathbb{E}_t[D_{i,t+h}]}{f_{i,d,t}^{(h)}}\right)^{1/h}-1. \tag{4}$$

Here, $r_{i,t+h}^h$ denotes the expected return on the strip paying firm i 's dividend at horizon $t+h$. Because $f_{i,d,t}^{(h)}$ is a futures price, this expected return measure is already in excess of the risk-free rate. The object is a yield-to-maturity rather than an expected one-period return. As emphasized by Gormsen and Lazarus (2023), a cleaner mapping to the cross-section of stock returns would be through expected one-period returns, but this is not feasible here because next-period expected strip prices are not observed. The expected yield-to-maturity is therefore the natural measure available in the data.

Based on these expected returns, we further compute expected CAPM alphas as

$$\alpha_{i,t+h}^h=\mathbb{E}_t\left[r_{i,t+h}^h\right]-\beta_{i,\text{maturity}}^h\times 5\%. \tag{5}$$

Here, $\beta_{i,\text{maturity}}^{h}$ is the maturity-specific beta-yield of the h -year strip, and 5% is the assumed market risk premium. Thus, the expected CAPM alpha measures the component of strip expected return not explained by its exposure to aggregate market risk. The construction of the maturity-specific beta-yield is deferred to the Appendix.

These strip-level measures identify the pricing of near-term firm-level cash flows directly. Because each strip is tied to a particular payout year, differences in strip expected returns across carbon-sorted firms can be interpreted as differences in the pricing of carbon risk at a given horizon. In other words, this approach allows us to hold maturity fixed while varying firms' carbon characteristics.

D.2. Construction of The Long-Term Asset

Dividend futures provide direct evidence only on the near end of the term structure. To shed light on longer horizons, we construct a residual long-term asset that captures all cash flows beyond the set of observed short-horizon dividend strips.

We use two complementary approaches. The first is price-based. Conceptually, the stock price can be decomposed into the sum of prices of short-horizon dividend strips and the residual long-term asset:

$$P_{i,t} = \sum_{h=1}^{H} P_{i,d,t}^{(h)} + LTA_{i,t}, \tag{6}$$

where $P_{i,t}$ is the stock price of firm i at time t, $P_{i,d,t}^{(h)}$ is the present value of the h -year dividend strip, and $LTA_{i,t}$ is the price of the long-term asset. Given observed stock prices and observed strip prices for maturities $h = 1, \dots, H$, the long-term asset is obtained as the residual claim to cash flows beyond the short-horizon strip window. This price-based approach is useful for constructing realized returns on the long-term asset. One should note, however, that the resulting long-term asset mechanically inherits SSDF prices, which may be stale. As a result, realized long-term asset returns may be driven disproportionately by stock-price movements rather than by variation in strip prices.

The second approach is return-based. From the decomposition in Section II.A, the expected return on a stock can be written as the weighted average of the expected returns on its future cash-flow components. Restricting attention to the observed short-horizon strips and the residual long-term asset gives

$$\mathbb{E}_t\left[r_{i,t+1}^{\text{Stock}}\right] = \sum_{h=1}^{H} \omega_{i,t}^{h}\ \mathbb{E}_t\left[r_{i,t+h}^{h}\right] + \left(1 - \sum\nolimits_{h=1}^{H} \omega_{i,t}^{h}\right) \mathbb{E}_t\left[r_{i,t+1}^{\text{LTA}}\right], \tag{7}$$

where $\omega_{i,t}^{h}$ is the ex ante fractional weight of the h -year strip in the stock, defined as $P_{i,d,t}^{(h)}/P_{i,t}$, and $\mathbb{E}_t[r_{i,t+1}^{\text{LTA}}]$ is the expected return on the long-term asset. Given the stock's expected return, the expected returns on the observed short-horizon strips, and their associated weights, equation (7) allows us to back out the expected return on the long-term asset.

In practice, the stock expected return used in this decomposition is estimated from rolling-window FF6 regressions. We estimate time-varying stock risk exposures and combine them with corresponding factor risk premia, measured as average factor returns over the rolling window, to obtain an estimate of the stock's expected return. This return-based approach is particularly useful because, as discussed above, SSDF prices are not well suited to high-frequency analysis.

Together, the price-based and return-based approaches provide a way to identify the far end of the cash-flow term structure. The first recovers the residual long-horizon claim from stock and strip prices, while the second infers its expected return from the stock's overall expected return and the expected returns on observed short-horizon strips.

E. Sample Coverage, Summary Statistics, and Representativeness

E.1. Sample Evolution and Maturity Coverage

This section describes the final matched sample used in the portfolio analysis. The sample starts in January 2012 and ends in December 2022. The start date reflects the limited cross-sectional coverage of the SSDF market in the early years, while the end date reflects the availability of carbon emissions data and the desire to avoid relying on excessively stale sorting variables after financial year 2020. As discussed above, the final sample contains 13,226 firm-month observations for 189 firms over the 2012–2022 period.

Figure 2 summarizes how the sample evolves over time and how maturity coverage is distributed. The upper panel shows that the number of firms in the matched sample rises steadily from 30 at the beginning of the sample to more than 150 by the end. The same panel also shows that the average number of maturity observations per firm remains limited throughout the sample and generally fluctuates around three. As discussed above, this partly reflects the fact that long-maturity SSDFs are excluded from the sample when their open interest is zero. This matters especially early in the life cycle of long-dated contracts, when listed products may exist but have not yet attracted unsettled positions in the market. Excluding such observations helps ensure that the prices used in the analysis reflect actual market valuation of future cash flows rather than purely mechanical quoted values.

Insert Figure 2 here

The lower panel of Figure 2 makes this point more directly. The sample is dominated by firms with one- to four-year strip observations, while the fraction of firms with five-year coverage remains modest throughout. In principle, most firms have listed annual contracts extending up to five years ahead, and for a selected subset even longer. In practice, however, five-year contracts are rarely active in the first year of their life cycle. For empirical purposes, the usable term structure is therefore concentrated at maturities of up to four years, and the subsequent dividend-strip analysis excludes contracts with maturity greater than or equal to five years.

E.2. Summary Statistics

Table 1 reports summary statistics for the matched portfolio sample. Panel A focuses on single-stock dividend-futures variables measured at calendar year-end, so that one- to four-year annual contracts correspond cleanly to maturities of 12, 24, 36, and 48 months. Several features stand out. First, the strip sample is concentrated at short maturities. One-year and two-year strips account for the largest shares of the sample, with maturity-dummy means of 0.37 and 0.33, respectively, followed by three-year strips at 0.22. Four-year strips are already relatively sparse, with a mean of 0.07. This maturity profile is remarkably close to that in Gormsen and Lazarus (2023), whose matched SSDF sample reports corresponding dummy means of 0.36, 0.33, 0.22, and 0.09, respectively, again showing that the data are concentrated in short-maturity contracts.

Insert Table 1 here

Second, trading activity and open interest vary substantially across observations. Annual trading volume has a mean of 10.24 thousand contracts but a median of only 1.83 thousand, while open interest has a mean of 6.81 thousand and a median of 1.65 thousand. Notional open interest shows a similar pattern. These large gaps between means and medians indicate that trading activity is concentrated in a relatively small subset of contracts, consistent with the substantial heterogeneity in liquidity discussed in Section II.C.

Third, the average expected return in the strip sample is economically sizeable. The mean annual expected return is 13.4%, while the median is 9.09%. These values are high relative to conventional estimates of stock-level expected returns and are consistent with the downward-sloping equity term structure documented in the aggregate market. Gormsen and Lazarus (2023) emphasize that risk-adjusted returns on cash flows decline with maturity and that near-future cash flows command especially high returns relative to conventional risk measures. Our evidence suggests that this pattern is also present at the individual-firm level.

Finally, the average dividend-strip-to-price ratio is small relative to the stock price, as expected for claims on short-horizon dividends only. The mean DS/P ratio is 3.71%, and the median is 3.66%. We further decompose this ratio by maturity. Because strip availability differs across maturities, we condition this comparison on the existence of the four-year strip so that the cross-maturity comparison is based on a common set of firm observations. Conditional on the four-year SSDF being available, the mean DS/P ratio declines from 4.31% for one-year strips to 4.10%, 3.95%, and 3.86% for the two-, three-, and four-year strips, respectively. This pattern indicates that more distant dividend claims account for progressively smaller shares of firm value, consistent with stronger discounting of longer-horizon cash flows.

Panel B of Table 1 shows that the carbon variables in the final sample display substantial cross-sectional variation. Scope 1 intensity is much more dispersed than Scope 2 intensity: the mean and standard deviation are 129.78 and 275.33 for Scope 1 intensity, versus 34.19 and 59.37 for Scope 2 intensity. This pattern is consistent with the broader carbon literature. In Bolton and Kacperczyk (2021), the corresponding means are roughly 192 and 34 in the same underlying units of tons of CO_2e per million U.S. dollars of sales. In other words, our intensity variables are of the same order of magnitude, and, as in their data, Scope 1 intensity is clearly more dispersed than Scope 2 intensity.

By contrast, the log-emissions measures are more tightly centered. Their means and medians are close to one another, and their standard deviations are modest relative to their levels. For Scope 1, the mean, median, and standard deviation are 12.96, 12.79, and 2.67, respectively; for Scope 2, they are 13.03, 12.99, and 1.63. At the same time, the level of emissions in our sample is much larger than in broader equity samples. Zhang (2025), using U.S. and global samples over 2009–2021, reports mean log emissions around 10 for both Scope 1 and Scope 2 in both the U.S. and global samples. Relative to those benchmarks, our mean log-emissions values of about 13 imply physical emissions that are roughly 20 times larger in levels. This is not surprising. The firms that enter the SSDF market are very large firms, so even at similar carbon intensities, their total emissions are naturally much larger.

Panel C of Table 1 further shows that the matched firms are large, established companies. The average log market capitalization is 10.64. As a benchmark, the minimum log market capitalization among S&P 500 constituents at the end of 2022 is approximately 8, which is below the minimum in our sample. This suggests that all firms in our sample would fall within the size range of the S&P 500. More generally, when we compute each firm's market-cap percentile within its domestic market, the average percentile is about 95.6, indicating that these firms belong to the largest listed firms in their respective countries. This is very much in line with Gormsen and Lazarus (2023), who report an average market-cap percentile of 97.2 for the firms underlying their SSDF sample from 2010 to 2019 and conclude that the main dimension along which the sample differs from the full universe is market size.

Taken together, Figure 2 and Panel A of Table 1 show that the final sample expands meaningfully over time and delivers a usable cross section for portfolio sorts, but that the pricing information remains concentrated at the front end of the term structure.

E.3. Representativeness and Data Quality

The evidence above points to an important feature of the sample: the firms that enter the SSDF market are typically large and important firms in their home and global markets. This feature is economically useful rather than problematic. Large firms naturally dominate value-weighted equity portfolios, and they are also the firms most exposed to scrutiny by investors, regulators, and the public during the transition toward net zero. In that sense, the part of the market most relevant for value-weighted carbon-return evidence overlaps closely with the firms that appear in the SSDF sample. The sample is therefore not a random slice of the equity market; it is the segment in which carbon transition risk is most likely to be salient and most likely to be reflected in asset prices.

In addition, these large firms tend to have better disclosure practices and greater transparency, so emissions data quality in this sample is materially stronger than in broad-market samples. A central concern in the carbon literature is that vendor-estimated emissions may mechanically reflect firm fundamentals, especially sales, rather than true environmental performance. Zhang (2025) emphasizes that emissions are tightly linked to firm activity and that vendor-estimated emissions often rely on industry averages and firm size or growth. Figure 3 suggests that these concerns are much less severe in the present sample. The left panel shows that 68.2% of the matched observations are directly disclosed and another 27.4% are derived from disclosures, leaving only 4.5% as purely estimated observations. The right panel shows that directly disclosed observations make up the majority of the monthly sample throughout most of the period, while purely estimated observations remain a small minority and are concentrated mainly in the earlier years. This composition contrasts sharply with the broad-sample concern highlighted by He, Li, and Zhang (2025). In their sample, estimated observations account for 11,059 out of 14,935 firm-year observations, or about 74.0% overall. They also show that estimated observations are especially prevalent among low-emission firms and can dominate the green end of the cross section. In our setting, by contrast, the dominance of disclosure-based observations substantially reduces the role of purely vendor-generated emissions.

Insert Figure 3 here

Finally, the sample is broadly distributed across sectors. Figure 4 shows that financials account for the largest share of firm-month observations, at 20.1% of the sample, but the remaining observations are spread fairly evenly across consumer staples (11.4%), health care (11.0%), consumer discretionary (10.8%), industrials (10.2%), utilities (9.0%), communication services (8.8%), materials (7.9%), energy (4.9%), and information technology (4.9%), with real estate accounting for only 0.9%. The right panel shows that this sector composition evolves over time but remains broadly diversified. This is important because it means the sample is not merely a fossil-fuel or heavy-industry sample. Instead, it covers a broad set of sectors that are central to both the modern economy and the climate-transition debate.

Insert Figure 4 here

Overall, the sample is concentrated in large developed-market firms, spans a broad set of sectors, relies primarily on disclosure-based emissions data, and provides meaningful short-horizon strip coverage. Although the sample is not intended to represent the full equity universe, it is in several respects especially well suited to the question in this paper. These features make it especially suitable for studying the term structure of carbon return.

III. Stock-Level Carbon Return Results

In this section, we examine whether a negative carbon return is already visible at the stock level in our sample. At the beginning of month *t*, firms are assigned to tercile portfolios using Scope 1 or Scope 2 carbon intensity measured at the end of month t−1, and portfolio returns over month t are then computed under both value-weighted and equal-weighted schemes. In the value-weighted specification, portfolio weights are proportional to firms' U.S.-dollar market capitalization, while the equal-weighted specification assigns the same weight to each stock within a portfolio. Because the sample is drawn primarily from developed markets, including Europe and the United States, risk adjustment is based on the developed-market Fama–French six-factor model. Specifically, for each portfolio *p*, I estimate a time-series regression

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_{MKT,p} MKT_t + \beta_{SMB,p} SMB_t + \beta_{HML,p} HML_t + \beta_{RMW,p} RMW_t + \beta_{CMA,p} CMA_t + \beta_{WML,p} WML_t + \varepsilon_{p,t}, \tag{8}$$

where $R_{p,t} - R_{f,t}$is the portfolio excess return in month *t*, and $\alpha_p$is the FF6-adjusted abnormal return. Newey–West *t*-statistics with three monthly lags are reported in parentheses. All non-U.S.-dollar stock prices and market capitalizations are converted into U.S. dollars using contemporaneous exchange rates prior to portfolio formation. The resulting stock-level portfolio evidence is reported in Table 2.

Insert Table 2 here

The value-weighted results in Panel A reveal a clear negative carbon return at the stock level. When firms are sorted on Scope 1 intensity, the alpha declines monotonically from 0.38% per month for the low-carbon portfolio to –0.18% for the high-carbon portfolio, implying a high-minus-low alpha of –0.55% per month (t = –2.84). The pattern is very similar for Scope 2 intensity, where the H–L alpha is –0.44% (t = –2.35). These are economically meaningful spreads, corresponding to roughly 6.6% and 5.3% per year, respectively, under simple annualization. They are also quantitatively close to Zhang (2025), who reports value-weighted H–L alphas of –0.40% and –0.34% for U.S. portfolios and –0.40% and –0.33% for developed-market portfolios when sorting on Scope 1 and Scope 2 intensity, respectively. Thus, despite relying on a much smaller SSDF-based universe, Table 2 recovers a stock-level carbon return that is broadly in line with the developed-market evidence in the prior literature.

Panel B shows that the result is not driven solely by a handful of mega-cap firms. Under equal weighting, the H–L alpha is –0.52% per month for both Scope 1 and Scope 2 intensity, with *t*-statistics of –2.71 and –2.79, respectively. The negative carbon return therefore survives when the portfolio is no longer dominated by the largest firms. This is notable because, in studies with much broader cross-sectional coverage, moving from value weighting to equal weighting often weakens statistical significance. Here, by contrast, the result remains strong under both weighting schemes, suggesting that the observed negative carbon return is not merely an artifact of capitalization weighting.

The factor loadings help clarify what this negative alpha is, and is not, capturing. Consider first the market factor. In Panel A, the market beta tends to decline as carbon intensity rises, from 0.84 to 0.77 for Scope 1 and from 0.80 to 0.75 for Scope 2. The pattern is present but not perfectly monotonic, and the H–L market loading is small and statistically insignificant. Under equal weighting, this difference becomes even weaker. This is consistent with the idea that the sample is already concentrated in large developed-market firms, all of which comove strongly with the developed-market market factor. As a result, there is limited remaining cross-sectional variation in market beta within this already large-cap subsample, even though the broader literature often finds somewhat higher market betas for greener firms.

By contrast, the size loading varies much more systematically. Across weighting schemes, the H–L SMB loading is significantly positive, and the pattern is more pronounced for Scope 1 than for Scope 2. This indicates that greener portfolios tilt more strongly toward large-cap firms, while browner portfolios, although still large in absolute terms, behave like relatively smaller firms within this large-cap universe. One possible interpretation is that carbon intensity, by scaling emissions with a size-related denominator such as sales, may mechanically induce a negative relation with firm size. Because Scope 1 emissions are more closely tied to firms' directly owned operations and production assets, this relation may be particularly strong for Scope 1 intensity.

The HML loading does not show a similarly clean pattern. Although the H–L HML loading is negative for both Scope 1 and Scope 2, the loadings are not uniformly monotonic across the three terciles.

The profitability and investment loadings move in economically sensible ways. In Panel A, the H–L RMW loading is positive, at 0.40 ($t = 1.72$) for Scope 1 and 0.56 ($t = 2.71$) for Scope 2, while the H–L CMA loading is strongly positive, at 0.93 ($t = 3.80$) and 0.73 ($t = 3.29$), respectively. The same pattern remains under equal weighting. These loadings are consistent with a plausible economic interpretation: browner firms may devote less effort to improving carbon efficiency, remain more conservative in investment, and therefore exhibit stronger profitability and investment-factor exposures, whereas greener firms appear more investment-intensive and have relatively weaker current profitability, potentially reflecting greater adoption of green technologies or transition-related expenditures.

Importantly, however, controlling for these factor exposures does not eliminate the return spread. The H–L alpha remains significantly negative after FF6 adjustment. The sample is not a random slice of the global equity market; it is concentrated in very large firms from developed markets, especially Europe, together with a smaller set of U.S. firms. These are exactly the firms that tend to receive the largest weights in aggregate developed-market portfolios. In that sense, even though the cross section is narrower, the sample still contains many of the firms most relevant for value-weighted carbon-return estimates in the broader literature. This helps explain why the value-weighted H–L alpha in Table 2 remains both economically large and statistically significant. It is also consistent with the broader finding that carbon returns are more negative in developed markets than in emerging markets.

Overall, Table 2 shows that even within the relatively narrow SSDF-linked stock universe, more carbon-intensive firms earn significantly lower future stock returns than less carbon-intensive firms. The result appears in both value-weighted and equal-weighted specifications, holds for both Scope 1 and Scope 2 intensity, survives FF6 adjustment, and is quantitatively close to the developed-market evidence in Zhang (2025). Table 2 therefore establishes an important starting point for the rest of the paper: the negative carbon return is already clearly visible at the stock level in this sample.

IV. Short-Term Dividend Strip Results

In this subsection, we examine how the carbon return is priced at the cash-flow level rather than at the stock level. The goal is twofold. First, holding carbon intensity fixed, we study how expected returns and CAPM alphas vary across dividend strips of different maturities, that is, the term structure of carbon-related cash-flow premia. Second, holding cash-flow maturity fixed, we examine how these expected returns and alphas vary across firms with different carbon-emission intensity. This design makes it possible to separately characterize how carbon-related pricing varies across both firm-level emissions exposure and cash-flow horizon.

To implement this exercise, we use annual single-stock dividend futures and retain only December observations. Because these contracts expire on the third Friday of December each year, restricting the sample to December ensures that the one-year to four-year strips correspond cleanly to claims with 12, 24, 36, and 48 calendar months to maturity, respectively. At a higher sampling frequency, strip maturity would no longer remain an integer number of years, and the maturity of otherwise comparable strips would vary mechanically from month to month. Restricting attention to December therefore allows maturity to be held fixed across observations. At each December month-end, dividend strips are assigned to tercile portfolios based on the underlying firm's Scope 1 or Scope 2 carbon intensity [8], and expected returns and CAPM alphas are measured separately for each maturity. Because expected returns are ex ante objects, they are observed at portfolio formation and do not require a subsequent holding period to be realized. We consider both value-weighted and equal-weighted portfolio constructions, where the value-weighted specification uses U.S.-dollar notional open interest, which can be viewed as an analogue to market capitalization in the SSDF setting. The resulting evidence is reported in Tables 3 and 4.

Insert Table 3 here

Table 3 reports expected returns on carbon-intensity-sorted dividend strip portfolios. Reading the table horizontally, and therefore holding carbon intensity fixed, expected returns generally decline with maturity under both weighting schemes and for both Scope 1 and Scope 2 intensity. Under value weighting, for example, the low-carbon Scope 1 portfolio declines from 12.75% at the one-year horizon to 9.03% at the four-year horizon, while the corresponding high-carbon portfolio declines from 9.21% to 8.05%. A similar downward-sloping term structure appears for Scope 2. The decline becomes less steep at longer maturities, but the overall pattern remains that short-maturity strips command higher expected returns than long-maturity strips. Under equal weighting, the same broad conclusion continues to hold across carbon-intensity groups. In this respect, the evidence is consistent

[8] *The sorting is carried out at the dividend-strip level rather than by first forming stock portfolios and then aggregating the strips within each stock group. This is important for two reasons. First, strip availability differs across firms and maturities, so sorting stocks first could mechanically distort portfolio composition if SSDF coverage is uneven across carbon-intensity groups. Second, strip-level sorting avoids having the results be driven by variation in the dividend-strip-to-price ratio, that is, by the cross-sectional distribution of cash-flow duration at the stock level. The exercise is therefore best interpreted as asking two separate questions: holding maturity fixed, how does the carbon return vary across emission-intensity groups; and holding emission intensity fixed, how is the carbon return priced across cash flows of different maturities.*

with Gormsen and Lazarus (2023), who show that cash-flow premia decline with maturity and that near-future cash flows command especially high returns.

More importantly, the vertical reading of Table 3 shows that the carbon return is strongly maturity-dependent. Holding cash-flow maturity fixed, brown firms earn substantially lower expected returns than green firms at short horizons. Under value weighting, the Scope 1 high-minus-low spread is –3.54% at the one-year horizon and –2.46% at the two-year horizon, both statistically significant, before shrinking to –1.90% at three years and an insignificant –0.98% at four years. Scope 2 displays the same short-horizon pattern: the H–L spread is significantly negative at the one- and two-year horizons, remains negative but smaller at three years, and then reverses sharply to 2.93% at the four-year horizon. This indicates that although all carbon-intensity groups exhibit downward-sloping term structures, the term structure of the green portfolios declines more steeply from near-term to long-term cash flows. As a result, the H–L portfolio itself exhibits a downward-sloping term structure: the carbon return is significantly negative at short maturities but converges rapidly toward zero as maturity increases. The equal-weighted results in Panel B deliver a similar message. For Scope 1, the H–L spread remains negative even at four years, but both its economic magnitude and its statistical significance tend to weaken as maturity lengthens. Taken together, the evidence in Table 3 shows that the negative carbon return is concentrated in short-maturity cash flows and fades quickly with horizon. In some specifications, especially for Scope 2, it even turns positive at the longest maturity of 4 years.

This maturity pattern is also visible in the monotonicity across the underlying carbon terciles. At short maturities, the ordering is generally clean: low-carbon portfolios earn the highest expected returns, high-carbon portfolios earn the lowest, and the middle portfolio lies between them. As maturity increases, however, this monotonic ordering becomes less systematic. The H–L spread shrinks, statistical significance weakens, and the ranking of the three terciles becomes less stable. This is exactly what one would expect if the negative carbon return loads primarily on near-term cash flows rather than applying uniformly across the entire cash-flow horizon. The main message of Table 3 is therefore not simply that brown firms underperform green firms on average, but that this underperformance is largely a short-horizon phenomenon concentrated in near-term dividend strips.

Insert Table 4 here

Table 4 sharpens this point by moving from raw expected returns to CAPM-adjusted alphas. Reading horizontally, CAPM alphas also tend to decline with maturity within each carbon portfolio, again consistent with a downward-sloping term structure in risk-adjusted cash-flow premia. The vertical reading, however, remains the more informative one. Under value weighting, the carbon-alpha spread is weak for Scope 1 at all maturities. For Scope 2, the spread is likewise small at short maturities but turns significantly positive at longer horizons, reaching 1.10% at the three-year horizon and 3.63% at the four-year horizon. [9] Under equal weighting, by contrast, the short-horizon negative carbon alpha closely mirrors the expected-return results in Table 3. For Scope 1, the H–L alpha is –5.39% at one year and –3.98% at

[9] *The value-weighted CAPM-alpha results exhibit weaker monotonicity across carbon-intensity groups, especially in the lowest-emission tercile. A plausible explanation is that notional open interest is more unevenly distributed within the greenest group, so that a subset of relatively liquid SSDF contracts receives disproportionately large weights. Since those contracts also have relatively high market betas and therefore relatively small CAPM alphas, the resulting weighted average will mechanically flatten the downward-sloping term structure of the green portfolio and weaken the short-horizon H–L spread. In unreported results, we also use firms' U.S.-dollar market capitalization as the weighting variable instead of notional open interest. Under that alternative weighting scheme, the negative one- to two-year H–L CAPM alpha regains statistical significance. These results are available upon request.*

two years, before moving much closer to zero and becoming statistically indistinguishable from zero at the three- and four-year horizons. For Scope 2, the equal-weighted H–L alpha is –3.67% at one year and –1.81% at two years, but disappears thereafter and becomes positive, though insignificant, by four years. Thus, once risk adjustment is imposed, the same economic conclusion continues to hold: the negative carbon premium is strongest in short-maturity strips and fades rapidly with maturity.

At first glance, these findings may appear to differ from Gormsen and Lazarus (2023). Their central result is that CAPM alphas on dividend strips decline with maturity within firms, while at a fixed maturity they do not vary systematically across firms for standard cross-sectional predictors such as value, profitability, investment, low-risk, and payout. This is not, however, a contradiction of their result. Rather, our evidence suggests that carbon intensity contains information that is especially relevant for the pricing of maturity-specific cash flows. Put differently, broad stock characteristics may be poor proxies for short-horizon equity premia in general, whereas carbon intensity appears to sort firms in a way that is particularly informative about the pricing of short-maturity dividend claims. In that sense, the difference itself is informative: it suggests that cash flows of different maturities may differ systematically in their exposure to carbon-related risk, which in turn underscores the importance of studying how carbon risk is priced across the cash-flow horizon rather than treating it as a single maturity-invariant object.

The main implication of Tables 3 and 4 is therefore about the term structure of the carbon return. Across both expected returns and CAPM alphas, the evidence points to a downward-sloping short-end structure: the carbon return is economically large and significantly negative at the one- and two-year horizons, becomes smaller at three years, and approaches zero—or in some specifications turns positive—by four years. The pattern is strongest and cleanest in expected returns, but it remains visible, though less uniformly, in CAPM alphas. Carbon risk should therefore not be modeled as a single maturity-invariant pricing effect. Instead, its pricing appears to depend crucially on cash-flow horizon, with the strongest negative premium concentrated in near-term claims.

V. Long-Term Asset Results

In this section, we turn to the far end of the cash-flow term structure by examining the residual long-term asset. The stock-level results in Section III show a negative carbon return for the stock as a whole, while Section IV shows that this negative carbon return is concentrated in short-maturity dividend strips and fades rapidly with maturity. The natural next question is therefore how carbon return is priced in the remaining long-horizon cash flows once the observable short-term strips have been removed. This is precisely the role of the long-term asset. It captures the residual claim to cash flows beyond the observed short-horizon dividend strips and therefore provides a complementary view of how carbon-related pricing behaves at the far end of the term structure.

To study this object, we sort long-term assets into tercile portfolios using the same firm-level Scope 1 and Scope 2 carbon intensity measures as before. Table 5 reports two sets of results. Panels A.1 and B.1 report annual expected returns using December observations only, mirroring the approach in Tables 3 and 4. Panels A.2 and B.2 report monthly realized-return evidence using the full monthly sample, where FF6 alphas are estimated from monthly regressions. Table 5 therefore allows the long-term asset to be compared directly with both the stock-level portfolios in Table 2 and the short-term strip portfolios in Tables 3 and 4.

Insert Table 5 here

The first striking result in Table 5 is that the sign of the carbon return reverses at the long end of the term structure when expected returns on the long-term asset are considered. In Panel A.1, the annual expected return on the low-carbon long-term asset portfolio is only 2.40% under Scope 1 sorting and 2.74% under Scope 2 sorting, whereas the high-carbon portfolio earns 7.78% and 8.49%, respectively. The resulting high-minus-low spread is 5.38% for Scope 1 and 5.75% for Scope 2, both highly statistically significant. The equal-weighted results in Panel B.1 lead to the same conclusion. Under equal weighting, the H–L expected-return spread is 6.73% for Scope 1 and 6.37% for Scope 2. Thus, unlike the stock as a whole, and unlike the short-horizon dividend strips, the long-term asset exhibits a positive carbon premium in expected returns: the distant cash flows of brown firms command higher expected returns than those of green firms.

This reversal is economically central. Section IV shows that the carbon return has a downward-sloping term structure at the front end: short-term dividend strips display a strongly negative carbon return, especially at the one- and two-year horizons, but both the economic magnitude and the statistical significance weaken with maturity and move toward zero. Table 5 shows that once attention shifts to the residual long-horizon claim, the sign flips. Taken together, these results imply that the term structure of carbon return is not merely downward-sloping toward zero, but appears to cross zero and become positive at sufficiently long maturities. In this sense, the stock-level negative carbon return in Table 2 should be understood as an aggregation outcome: the stock combines sharply negative carbon premia at the short end with positive carbon premia at the far end, and the overall stock-level effect reflects the balance of these horizon-specific components. This interpretation follows directly from the cash-flow decomposition in Section II, under which stock expected returns aggregate the expected returns on cash flows of different maturities using their ex ante value weights.

A second feature of the expected-return results is that, under both weighting schemes, the most robust pattern is a clear separation between the low-carbon portfolio and the other two terciles. The low-carbon portfolio has substantially lower expected returns than both the medium- and high-emission intensity portfolios. By contrast, the difference between the medium and high portfolios is comparatively small, especially under Scope 1 sorting, where the medium portfolio sometimes slightly exceeds the high portfolio in expected return. The main expected-return result in Table 5 is therefore not a perfectly monotonic increase across all three terciles, but rather a pronounced gap between the greenest tercile and the rest of the cross section.

The monthly realized-return evidence in Table 5 is more muted. In Panel A.2, the value-weighted FF6 alpha of the H–L long-term asset portfolio is 0.02 under Scope 1 sorting and –0.49 under Scope 2 sorting, neither of which is statistically distinguishable from zero at conventional levels. Panel B.2 conveys the same basic message under equal weighting: the H–L alpha is 0.00 for Scope 1 and –0.70 for Scope 2, again statistically weak. Thus, unlike the expected-return evidence, the realized-return regressions do not deliver comparably strong evidence of a positive carbon alpha for the long-term asset.

This weaker realized-return evidence should not be overstated in either direction. It does not overturn the expected-return results, but it does suggest that evidence from the far end of the term structure is noisier when measured using realized monthly returns. This is not surprising given how the long-term asset is constructed. As discussed in Section II, the realized long-term asset is a residual object backed out from stock prices and observed strip prices, and its price mechanically inherits noise from SSDF prices, which themselves can be stale. For this reason, the expected-return evidence in Panels A.1 and B.1 is more informative about the pricing of the long-term asset than the realized monthly FF6 alpha estimates in Panels A.2 and B.2.

Even so, the realized factor loadings in Panel A.2 remain informative about the economic nature of the long-term asset portfolios. The H–L long-term asset portfolio loads significantly negatively on the market factor under both Scope 1 and Scope 2 sorting, with coefficients of –0.21 and –0.23, respectively. This pattern is notably stronger than in the stock portfolios. In Table 2, market beta shows only a mild downward tendency from low- to high-emission portfolios, and the H–L spread is not statistically significant. By contrast, once the short-horizon dividend strips are removed, the long-term cash flows within each stock exhibit economically and statistically meaningful cross-sectional variation in market beta across carbon-intensity groups. In other words, the relation between higher emissions intensity and lower market beta becomes much sharper at the far end of the cash-flow term structure. This pattern is also broadly consistent with prior findings in the carbon-return literature, which often document an association between higher carbon emissions and lower market beta.

The long-term asset also loads strongly negatively on HML and strongly positively on CMA. These patterns are already visible in the stock portfolios in Table 2, but they become even stronger for the long-term asset. This suggests that the cross-sectional differences associated with carbon intensity are more sharply expressed once attention is restricted to long-horizon cash flows. At the long-horizon cash-flow level, greener firms resemble assets with stronger exposure to long-duration, growth-like cash flows, whereas browner firms resemble assets that are relatively more value-like and more conservatively invested. This interpretation is consistent with the broader picture emerging from the earlier sections: green firms appear to be associated with more aggressive investment and a larger share of value residing in distant cash flows, while brown firms' valuation is tilted relatively more toward nearer cash-flow structures.

Another notable feature concerns profitability exposure. At the stock level, the H–L portfolio exhibits a significantly positive loading on the profitability factor. One interpretation is that high-emission firms, by devoting less effort and fewer resources to emissions abatement and related transition activities, are able to sustain stronger current profitability. Interestingly, however, once the short-term cash flows are removed, this profitability exposure becomes statistically insignificant in the long-term asset. This suggests that the profitability advantage of brown firms relative to green firms is concentrated primarily in short-horizon cash flows. That interpretation is also economically intuitive. Any profitability disadvantage of green firms relative to brown firms is likely to be transitional rather than permanent. Over longer horizons, as net-zero commitments become more binding and green technologies are more widely adopted, green firms may reasonably be expected to perform better from a profitability perspective.

The most important implication of Table 5 is therefore that the carbon return changes sign across the term structure. Short-term dividend strips carry a negative carbon return, while the residual long-term asset carries a positive carbon return in expected returns. This is exactly the kind of pattern one would expect if carbon-related risk is priced differently across horizons rather than uniformly across the entire stock. More precisely, the evidence suggests that brown firms are penalized more heavily in their near-term cash flows, but command higher expected returns on their distant cash flows. Once these long-horizon components are isolated, the positive far-end carbon premium becomes visible.

Viewed together, Tables 2 through 5 paint a coherent picture. The stock-level negative carbon return in Table 2 reflects the pricing of all future cash flows combined. Tables 3 and 4 show that this negative premium is concentrated in short-maturity dividend strips and weakens rapidly with maturity. Table 5 then shows that at the far end of the term structure, the sign reverses: the long-term asset exhibits a positive carbon return in expected returns. The overall conclusion is that carbon return is fundamentally horizon-dependent. It is not a single maturity-invariant pricing effect, but a term-structure object whose sign and magnitude vary across cash-flow horizons. This is the central message of the paper.

VI. Robustness Analysis

This section examines whether the main results survive when the sorting variable is changed from carbon intensity to carbon emissions levels. Specifically, portfolios are now formed on log Scope 1 emissions and log Scope 2 emissions rather than emissions scaled by sales. This is a useful robustness check because emissions levels are a direct measure of firms' absolute carbon footprint, and some studies continue to argue that total emissions are the more relevant proxy for carbon transition risk. If the earlier results merely reflect the scaling embedded in emissions intensity, they should weaken materially once the sorting variable is replaced by emissions levels. Table 6 suggests otherwise. The main term-structure pattern remains broadly intact, especially for Scope 1 emissions, although the results are somewhat less uniform for Scope 2.

Insert Table 6 here

At the stock level, the negative carbon return remains visible when sorting on emissions levels. Under value weighting, the stock-level H–L alpha is a significant –0.45% for log Scope 1 emissions, while the corresponding spread for log Scope 2 emissions is also negative, though no longer significant. Under equal weighting, the same pattern remains, but statistical significance weakens. Thus, replacing emissions intensity with emissions levels does not overturn the stock-level evidence, but it does suggest that Scope 1 emissions levels contain cleaner cross-sectional pricing information than Scope 2 emissions levels at the stock level.

The short-term dividend strip results tell a similar story. For expected returns, the high-minus-low spread remains strongly negative at the short end and moves toward zero as maturity increases. The main result from Tables 3 and 4 therefore survives this alternative sorting scheme: the negative carbon return remains concentrated in near-term cash flows rather than applying uniformly across maturities.

For expected CAPM alphas on dividend strips, however, the robustness is more mixed. Under log Scope 1 emissions, the dividend-strip H–L CAPM alpha is close to zero under value weighting and negative only at the short end under equal weighting. Under log Scope 2 emissions, the value-weighted H–L alpha is in fact positive at short maturities, while the equal-weighted results are negative at the one- and two-year horizons but again weaken rapidly thereafter. Thus, the alpha evidence is less clean than the expected-return evidence when sorting on emissions levels, particularly for Scope 2.

Finally, the long-term asset results remain consistent with the sign-reversal argument in Table 5. When sorting on emissions levels, the long-term asset continues to exhibit a strongly positive H–L spread in annual expected returns, while the realized monthly alpha remains statistically weak. Thus, changing the sorting variable from intensity to level does not alter the broad message at the far end of the term structure: brown firms' distant cash flows continue to command higher expected returns than green firms' distant cash flows, even though the realized-return evidence remains noisy.

Overall, Table 6 shows that the main results are not driven by the choice of emissions intensity as the sorting variable. Sorting on emissions levels yields the same broad term-structure pattern: a negative carbon return at the stock level, a strongly negative carbon return concentrated in short-maturity dividend strips, and a positive carbon return in the long-term asset. The evidence is strongest for log Scope 1 emissions and somewhat weaker for log Scope 2 emissions, especially in CAPM alphas. Still, the central message survives: carbon return remains fundamentally horizon-dependent.

VII. Conclusion

This paper studies the term structure of carbon return. Rather than treating the stock as a single claim, we decompose it into short-horizon dividend strips and a residual long-term asset, and ask how carbon-related pricing varies across cash-flow maturities. Using single-stock dividend futures and analyst dividend expectations, we document that the stock-level negative carbon return in the SSDF-linked sample masks substantial horizon-specific heterogeneity.

The empirical results yield a clear picture. At the stock level, more carbon-intensive firms earn lower future returns. At the cash-flow level, this negative carbon return is concentrated in short-maturity dividend strips: it is economically large and statistically significant at the one- and two-year horizons, weakens rapidly with maturity, and in some specifications disappears or turns positive by four years. At the far end of the term structure, the sign reverses: the residual long-term asset exhibits a positive carbon return in expected returns. The overall stock-level carbon return is therefore an aggregation of negative short-end and positive long-end premia. Carbon return is not a single maturity-invariant object, but a horizon-dependent pricing pattern across firm cash flows.

These findings have two broader implications. First, they suggest that the climate-finance literature should move beyond stock-level average returns when studying the pricing of carbon risk. Once cash flows are separated by horizon, the pricing pattern is much sharper and more informative. Second, they show that cash-flow maturity is not merely a control variable, but a central dimension along which carbon-related risk is priced. Any model that seeks to explain carbon return should therefore allow for horizon-specific pricing effects rather than imposing a single carbon premium on the stock as a whole. More generally, the paper suggests that climate transition risk is best understood not only as a firm characteristic, but also as a term-structure object.

References

1. Alekseev, Georgij, Stefano Giglio, Quinn Maingi, Julia Selgrad, and Johannes Stroebel, 2022, A quantity-based approach to constructing climate risk hedge portfolios, Technical report, National Bureau of Economic Research.
2. Andrews, Spencer, and Andrei S. Gonçalves, 2020, The bond, equity, and real estate term structures, Working paper.
3. Ardia, David, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht, 2023, Climate change concerns and the performance of green vs. brown stocks, Management Science 69, 7607–7632.
4. Aswani, Jitendra, Aneesh Raghunandan, and Shiva Rajgopal, 2024, Are carbon emissions associated with stock returns, Review of Finance 28, 75–106.
5. Bolton, Patrick, and Marcin Kacperczyk, 2021, Do investors care about carbon risk?, Journal of Financial Economics 142, 517–549.
6. Bolton, Patrick, and Marcin Kacperczyk, 2023, Global pricing of carbon-transition risk, Journal of Finance 78, 3677–3754.
7. Cejnek, Georg, and Otto Randl, 2020, Dividend risk premia, Journal of Financial and Quantitative Analysis 55, 1199–1242.
8. Choi, Darwin, Zhenyu Gao, and Wenxi Jiang, 2020, Attention to global warming, Review of Financial Studies 33, 1112–1145.
9. Croce, Mariano M., Martin Lettau, and Sydney C. Ludvigson, 2009, Investor information, long-run risk, and the duration of risky cash-flows, Unpublished.
10. Da, Zhi, 2009, Cash flow, consumption risk, and the cross-section of stock returns, Journal of Finance 64, 923–956.
11. Dechow, Patricia M., Richard G. Sloan, and Mark T. Soliman, 2004, Implied equity duration: A new measure of equity risk, Review of Accounting Studies 9, 197–228.
12. Dyck, Alexander, Karl V. Lins, Lukas Roth, and Hannes F. Wagner, 2019, Do institutional investors drive corporate social responsibility? International evidence, Journal of Financial Economics 131, 693–714.
13. Gibson Brandon, Rajna, Simon Glossner, Philipp Krueger, Pedro Matos, and Tom Steffen, 2022, Do responsible investors invest responsibly? Review of Finance 26, 1389–1432.
14. Gonçalves, Andrei S., 2021, The short duration premium, Journal of Financial Economics 141, 919–945.
15. Gordon, Myron J., 1962, The Investment, Financing, and Valuation of the Corporation (Irwin, Homewood, IL).
16. Gormsen, Niels Joachim, 2021, Time variation of the equity term structure, Journal of Finance 76, 1959–1999.

17. Gormsen, Niels Joachim, and Eben Lazarus, 2023, Duration-driven returns, Journal of Finance 78, 1393–1447.
18. Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 1269–1286.
19. Hansen, Lars Peter, and Kenneth J. Singleton, 1983, Stochastic consumption, risk aversion, and the temporal behavior of asset returns, Journal of Political Economy 91, 249–265.
20. He, Yongqi, Jiangyuan Li, and Ruishen Zhang, 2024, Duration-driven carbon premium, Working paper.
21. Krueger, Philipp, Zacharias Sautner, and Laura T. Starks, 2020, The importance of climate risks for institutional investors, Review of Financial Studies 33, 1067–1111.
22. Lettau, Martin, and Jessica A. Wachter, 2007, Why is long-horizon equity less risky? A duration-based explanation of the value premium, Journal of Finance 62, 55–92.
23. Lucas, Robert E., Jr., 1978, Asset prices in an exchange economy, Econometrica 46, 1429–1445.
24. Mehra, Rajnish, and Edward C. Prescott, 1985, The equity premium puzzle, Journal of Monetary Economics 15, 145–161.
25. Miller, Shane, 2020, The term structures of equity risk premia in the cross-section of equities, Duke University Working Paper.
26. Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2021, Sustainable investing in equilibrium, Journal of Financial Economics 142, 550–571.
27. Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2022, Dissecting green returns, Journal of Financial Economics 146, 403–424.
28. Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, 2021, Responsible investing: The ESG-efficient frontier, Journal of Financial Economics 142, 572–597.
29. van Binsbergen, Jules H., and Ralph S. J. Koijen, 2017, The term structure of returns: Facts and theory, Journal of Financial Economics 124, 1–21.
30. van Binsbergen, Jules H., Michael Brandt, and Ralph Koijen, 2012, On the timing and pricing of dividends, American Economic Review 102, 1596–1618.
31. van Binsbergen, Jules H., Wouter Hueskes, Ralph Koijen, and Evert Vrugt, 2013, Equity yields, Journal of Financial Economics 110, 503–519.
32. Weber, Michael, 2018, Cash flow duration and the term structure of equity returns, Journal of Financial Economics 128, 486–503.
33. Zhang, Shaojun, 2025, Carbon returns across the globe, Journal of Finance 80, 615–644.

Appendix: Details on Single-Stock Dividend Futures

A. Product Mapping and Firm-Level Aggregation

The raw Eurex SSDF data are organized by the Product ID. Ideally, for a given firm, date, and payout year, one would observe a single contract. In practice, however, a firm can be associated with multiple product IDs for the same expiration. This occurs for three main reasons.

First, a standard SSDF contract is written on the dividends paid on 1,000 shares, but corporate actions can change the number of shares represented by a legacy contract. Stock splits, reverse splits, bonus issues, capital increases, repurchases, and related events can alter the share basis of the contract. In such cases, Eurex applies an adjustment factor to the legacy contract so that its economic value remains continuous. In practice, Eurex applies a published adjustment factor to the old contract, which changes the contract size and the quoted price correspondingly, and may simultaneously introduce a new standard contract written on 1,000 post-event shares. Old and new products can therefore coexist temporarily for the same firm and payout year.

Second, after events such as spin-offs, a legacy contract can cease to be a pure single-stock claim and instead become a basket dividend future, while a new standard product is introduced for the post-event firm. This again creates a period in which multiple product IDs coexist , although one contract is now a basket dividend future and the other remains a single-stock dividend future.

Third, a small number of firms have parallel contracts listed in different currencies.

We use Deutsche Börse header information to map each Product ID to the underlying firm ISIN and then to Compustat GVKEY. Before aggregation, we discard all observations with zero open interest and this removes listed but economically inactive contracts.

We then aggregate the data to the firm-date-expiration level. If a firm-date-expiration cell is associated with a unique product ID, the observation is retained as is. If multiple products arise only because of parallel currency listing, we retain the most actively traded version and discard the redundant contract. If several same-currency products map to the same firm-date-expiration cell and carry the same economic price, we treat them as parallel lines on the same underlying claim. In that case, trading volume and notional open interest are aggregated across products, while open interest is standardized to a common 1,000-share basis. If several same-currency products map to the same firm-date-expiration cell but carry materially different prices, we retain the latest product ID. These cases usually reflect periods in which an adjusted legacy contract coexists with a newer clean contract after a corporate event. In practice, this rule typically preserves the newer standard contract written on 1,000 shares.

In the raw panel, 1,448,221 contract-date-expiration observations are consolidated into 1,324,007 firm-date-expiration observations. The resulting firm-level panel is the input for all subsequent price and return construction.

B. Settlement Prices, Traded Prices, and Historical Price Conventions

The raw Eurex data report daily settlement prices. These are the end-of-day prices against which outstanding positions are cleared in the exchange's risk-management system. Settlement prices need not coincide with traded prices, because on no-trade days they can partly reflect quotes or exchange-side models. We therefore first construct a traded-price series in calendar time. For each firm-expiration claim, we update the price only on dates with strictly positive traded volume and otherwise carry the last traded price forward. This procedure prevents model-imputed settlement revisions from being interpreted as tradable price changes, especially for illiquid long-dated contracts.

When querying historical price series from Bloomberg, we work with two price conventions. Bloomberg provides three relevant adjustment flags: CapChg, which adjusts for capital changes such as stock splits, spin-offs, and bonus issues; CshAdjNormal, which adjusts for ordinary cash events such as regular dividends; and CshAdjAbnormal, which adjusts for abnormal cash events such as special dividends.

We construct an adjusted price series by setting CapChg = 1, CshAdjNormal = 0, and CshAdjAbnormal = 1. This series adjusts for capital changes and abnormal cash events, but not for ordinary dividends. It is therefore the appropriate series for time-series applications in which one wants to remove mechanical jumps caused by contract redesign while preserving the ordinary dividend stream that the contract is meant to track. In the empirical analysis, this adjusted series is used where time-series continuity is required, most importantly when constructing realized returns and estimating strip betas.

We also construct an unadjusted price series by setting CapChg = 0, CshAdjNormal = 0, and CshAdjAbnormal = 0. This series preserves the raw point-in-time contract price. It is the appropriate series for expected-return construction, because analyst dividend expectations are also point-in-time objects stated on a per-share basis. The relevant comparison is therefore between contemporaneous expected dividends and contemporaneous unadjusted strip prices.

In short, traded prices are the starting point for the return objects, while the distinction between adjusted and unadjusted Bloomberg histories determines whether the price series is suited to time-series continuity or to point-in-time matching with analyst expectations.

C. Return, Beta, and Alpha Construction

C.1. Expected and Realized Strip Returns

Expected returns are constructed by combining point-in-time analyst dividend expectations with unadjusted strip prices. For firm i, let $f_{i,d,t}^{(h)}$ denote the time-t futures price of the claim to the dividend paid at horizon $t+h$, and let $\mathbb{E}_t[D_{i,t+h}]$ denote the corresponding expected dividend. We define the expected return on the h -year strip as the expected yield-to-maturity,

$$\mathbb{E}_t\left[r_{i,t+h}^{h}\right] = \left(\frac{\mathbb{E}_t[D_{i,t+h}]}{f_{i,d,t}^{(h)}}\right)^{1/h} - 1. \tag{C1}$$

Because $f_{i,d,t}^{(h)}$ is a futures price, the return in (C1) is already measured in excess of the risk-free rate.

Realized returns are constructed from changes in futures prices over the holding period. For a claim with remaining maturity h at time t, the one-period realized return is

$$r_{i,t+1}^{h} = \frac{f_{i,d,t+1}^{(h-1)}}{f_{i,d,t}^{(h)}} - 1. \tag{C2}$$

At time t, we use traded prices. At time $t+1$, we again use traded prices unless the contract matures over the holding period, in which case the settlement price is used for the terminal leg. Because these are futures returns, the realized returns in (C2) are also measured in excess of the risk-free rate. Monthly realized returns constructed in the same way are used later in the estimation of strip betas.

C.2. Strip Betas and Beta-Yields

We estimate strip CAPM betas from monthly realized strip returns with lagged market returns to account for stale prices, following Dimson (1979) and Lewellen and Nagel (2006):

$$r_{i,t,t+1}^{h} = \beta_{0,i}^{h} + \beta_{1,i}^{h} r_{c,t+1}^{M} + \beta_{2,i}^{h} r_{c,t}^{M} + \beta_{3,i}^{h}\left(r_{c,t-1}^{M} + r_{c,t-2}^{M} + r_{c,t-3}^{M}\right) + \varepsilon_{i,t+1}^{h}, \tag{C3}$$

where $r_{c,t+1}^{M}$ is the excess return on the stock market in the country in which firm i is primarily traded. The strip beta is then

$$\beta_{i}^{h} = \beta_{1,i}^{h} + \beta_{2,i}^{h} + \beta_{3,i}^{h}. \tag{C4}$$

Because the regressions are monthly while maturity is measured in years, the remaining maturity of a claim is rounded up to the nearest integer. Thus, a claim is assigned to maturity bucket h when $12(h-1) <$ maturity in months $\leq 12h$. Following Gormsen and Lazarus (2023), we truncate strip betas to the interval $[-1, 1.5]$.

Because the expected-return object in (C1) is a yield-to-maturity rather than a one-period return, we construct a maturity-specific beta-yield as

$$\beta_{i,\text{maturity}}^{h} = \frac{1}{h}\sum_{j=1}^{h} \beta_{i}^{j}. \tag{C5}$$

C.3. Expected CAPM Alphas

Expected CAPM alphas are defined as

$$\alpha_{i,t+h}^{h} = \mathbb{E}_{t}\left[r_{i,t+h}^{h}\right] - \beta_{i,\text{maturity}}^{h} \times 5\%, \tag{C6}$$

where 5% is the assumed market risk premium used throughout the paper.

Figure 1. SSDF Market Expansion and Product Availability

This figure plots the evolution of the single-stock dividend futures market. The left axis reports the number of underlying firms, and the right axis reports the average number of December contracts per firm.

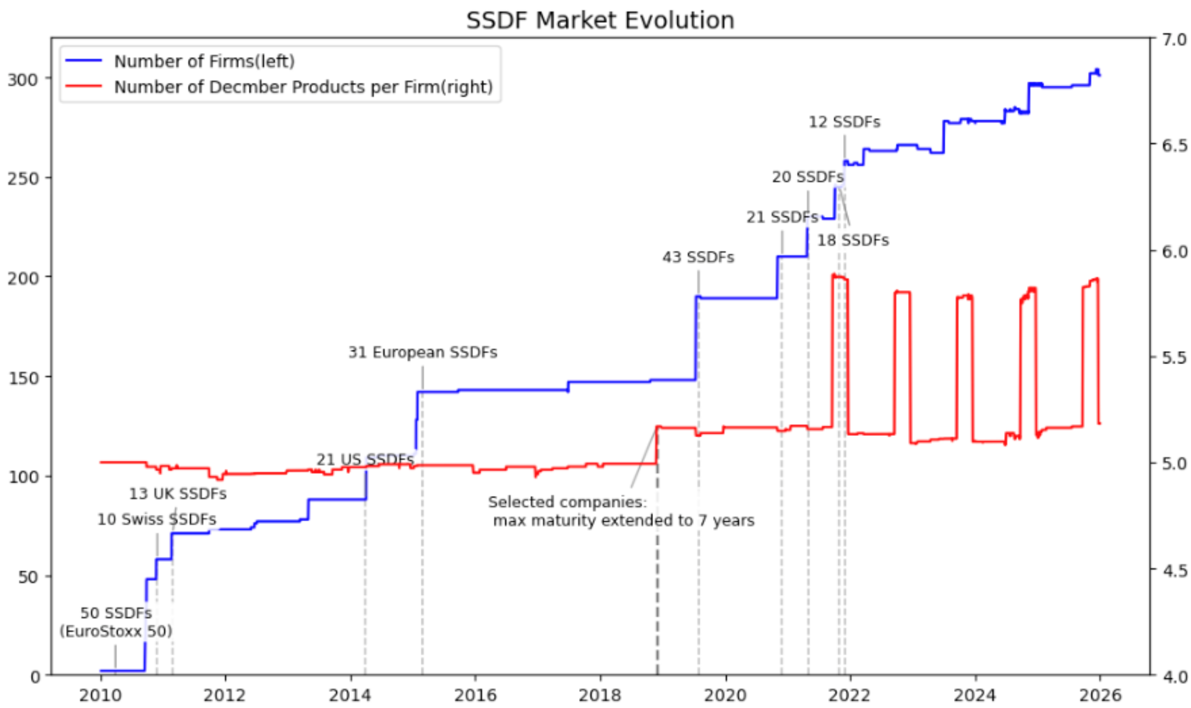


Figure 2. Sample Evolution and Maturity Coverage

This figure reports how the matched sample evolves over time and how maturity coverage is distributed. The upper panel shows the number of firms and the average number of maturity observations per firm. The lower panel shows the share of firms with available strip observations by maturity.

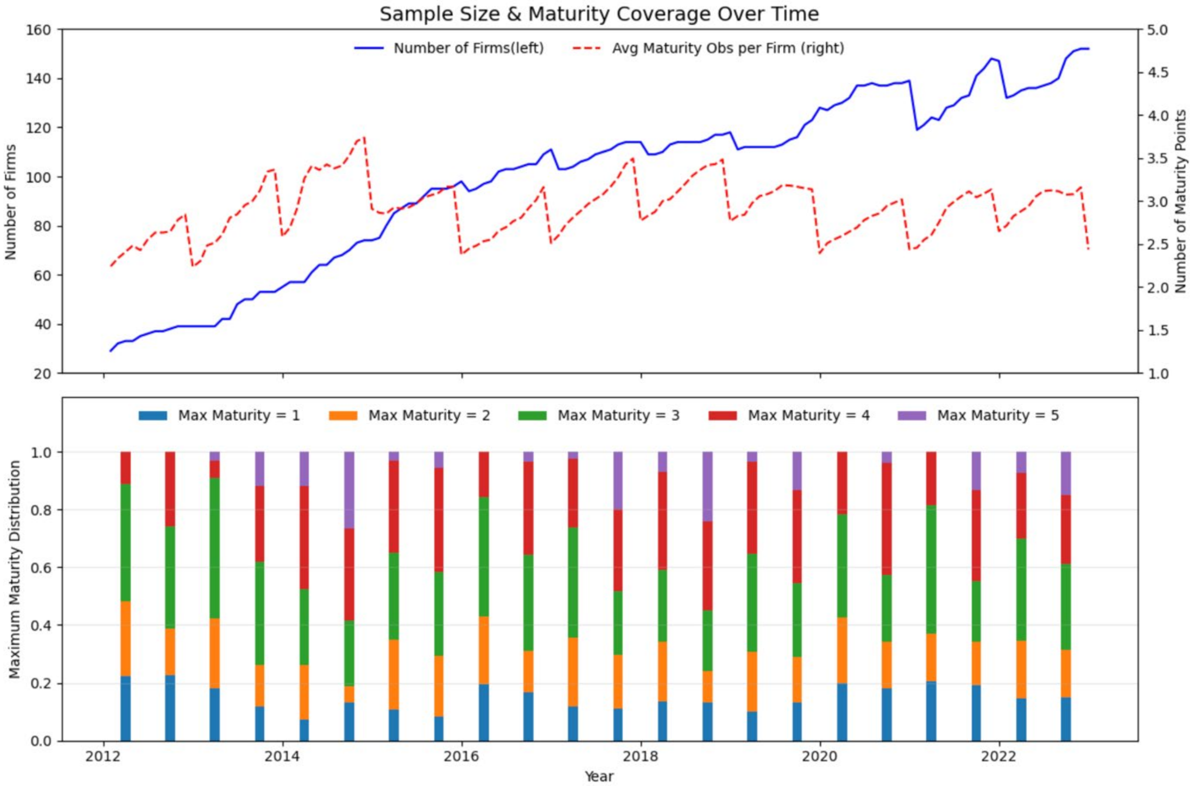


Figure 3. Disclosure Composition of Emissions Observations

This figure reports the disclosure status of emissions observations in the matched sample. The left panel shows the overall composition of disclosure types, and the right panel shows the composition over time.

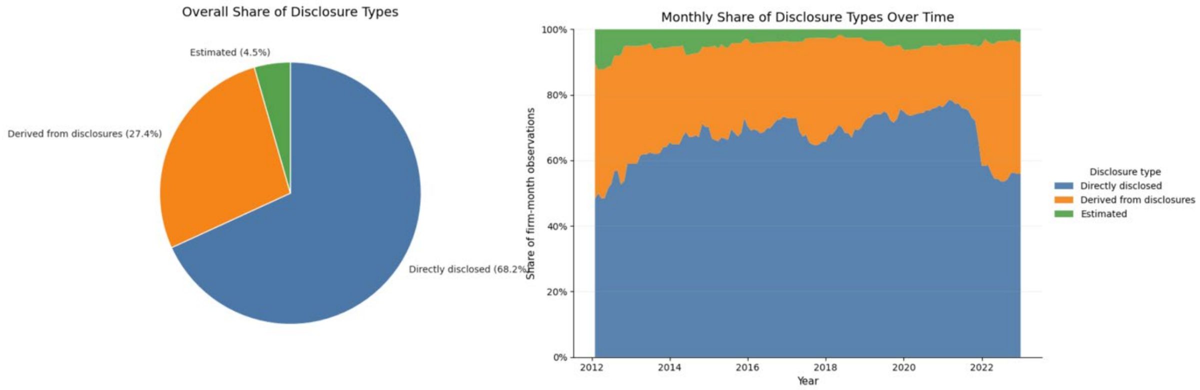


Figure 4. Sector Composition of the Matched Sample

This figure reports the sector composition of the matched sample. The left panel shows the full-sample distribution of firm-month observations across sectors, and the right panel shows how the sector composition evolves over time.

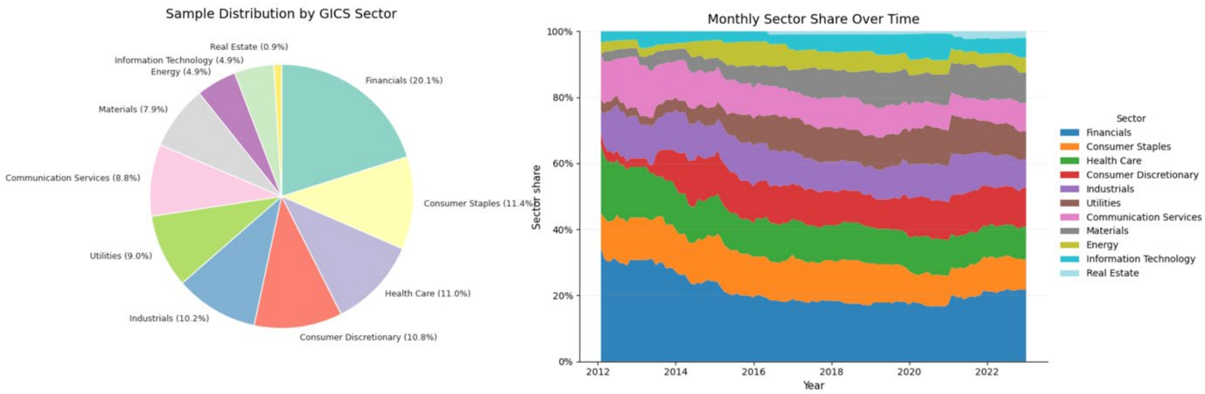


Table 1. Summary Statistics

This table reports summary statistics(number of observations, mean, standard deviation, minimum, median, and maximum) for the matched sample used in the portfolio analysis. The sample period is 2012 to 2022. At each month-end, the firm-month panel is updated using the most recently available annual carbon emissions data and annual financial statement information. For all non-dummy variables, we apply two-sided winsorization at the 2.5th and 97.5th percentiles. Panel A reports calendar-year-end single-stock dividend-futures variables. Because dividend strips with maturities of five years or longer are rare in the sample, we retain only strips with one- to four-year maturities. Annual trading volume is the total number of contracts traded over the calendar year. Open interest is the number of unsettled futures contracts at the end of the calendar year. Notional open interest is defined as open interest multiplied by contract size, typically 1,000 shares, and the dividend-futures settlement price per share. Dividend futures prices are converted into U.S. dollars using the corresponding exchange rate. To capture maturity coverage, we also report maturity dummies, whose means can be interpreted as the sample shares of dividend strips at different maturities. DS/P denotes the ratio of dividend-strip price to stock price. Expected returns are computed from unadjusted point-in-time traded prices and analyst expectations. Panel B reports six carbon-emissions measures based on the most recently released firm-financial-year observations that can be matched to the firm-month panel. Panel C reports selected firm characteristics constructed from the latest available annual report that can be matched to the firm-month panel. All accounting variables are first translated into U.S. dollars using the exchange rate corresponding to the fiscal year-end before the relevant variables are constructed.

Variable	Obs.	Mean	SD	Min	Median	Max
Panel A: Single Stock Dividend Futures						
Annual Expected Return (%)	2199	13.4	15.34	-7.83	9.09	74.49
DS/P (%)	2208	3.71	1.49	0.94	3.66	7.05
Open Interest (thousands)	2208	6.81	12.18	0.02	1.65	56.73
Notional Open Interest (USD millions)	2208	6.34	9.46	0.02	2.09	39.97
Annual Trading Volume (thousands)	2208	10.24	19.81	0	1.83	93.64
One-year maturity Dummy	2208	0.37	0.48	0	0	1
Two-year maturity Dummy	2208	0.33	0.47	0	0	1
Three-year maturity Dummy	2208	0.22	0.41	0	0	1
Four-year maturity Dummy	2208	0.07	0.26	0	0	1

Table 1. Summary Statistics (cont'd)

Variable	Obs.	Mean	SD	Min	Median	Max
Panel B: Carbon Emissions						
Carbon Intensity Scope 1 (tons CO2e/USD m.)	941	129.78	275.33	0.2	10.83	1195.93
Carbon Intensity Scope 2 (tons CO2e/USD m.)	941	34.19	59.37	0.57	14.09	288.87
Log (Carbon Emissions Scope 1 (tons CO2e))	939	12.96	2.67	8.34	12.79	18.13
Log (Carbon Emissions Scope 2 (tons CO2e))	941	13.03	1.63	9.73	12.99	16.12
Scope 1 Carbon Emissions Growth Rate (%)	937	-1.98	19.28	-43.2	-3.13	62.34
Scope 2 Carbon Emissions Growth Rate (%)	940	1.77	33.51	-53.51	-3.25	141.17
Panel C: Firm Characteristics						
Log (Market Cap (USD millions))	1230	10.64	0.92	8.75	10.61	12.53
Book-to-Market	1230	0.8	0.61	0.08	0.63	2.68
Tobin's Q	1230	1.51	0.84	0.77	1.17	4.55
ROA (%)	1254	9.84	6.69	0.73	9.34	27.45
Asset Growth Rate (%)	1254	2.83	10.53	-16.53	1.4	34.42
Leverage (%)	1260	69.09	17.77	34.11	68.2	96.31

Table 2. Carbon Intensity Sorted Stock Portfolios

This table reports monthly stock portfolio results sorted on carbon intensity. At the beginning of month t, stocks are assigned to tercile portfolios based on Scope 1 intensity or Scope 2 intensity available at the end of month t-1. Portfolio returns are then measured using realized stock returns over month t. Panel A reports value-weighted portfolio results, where returns are weighted by U.S.-dollar market capitalization. Panel B reports the corresponding equal-weighted portfolio results. For each portfolio, alpha and factor loadings are estimated from monthly regressions on the Fama-French six factors (Fama and French, 2018). Newey-West-adjusted t-statistics with three monthly lags are reported in parentheses below the coefficient estimates. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The sample period is 2012 to 2022.

Panel A: Value-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Alpha	0.38*	0.09	-0.18	-0.55***	0.26	0.24*	-0.18	-0.44**
	(1.95)	(0.71)	(-1.12)	(-2.84)	(1.55)	(1.66)	(-1.26)	(-2.35)
MKT	0.84***	0.72***	0.77***	-0.06	0.80***	0.77***	0.75***	-0.05
	(13.93)	(15.97)	(19.06)	(-1.07)	(15.11)	(16.77)	(16.94)	(-1.18)
SMB	-0.76***	-0.46***	-0.38***	0.38**	-0.64***	-0.53***	-0.41***	0.24*
	(-4.31)	(-3.35)	(-3.69)	(2.32)	(-4.69)	(-3.84)	(-3.33)	(1.71)
HML	0.43**	-0.32**	0.06	-0.37**	0.18	0.09	-0.15	-0.33**
	(2.42)	(-2.57)	(0.57)	(-2.06)	(1.22)	(0.60)	(-1.27)	(-2.04)
RMW	-0.32	0.22	0.08	0.40*	-0.31*	0.10	0.26	0.56***
	(-1.58)	(1.30)	(0.52)	(1.72)	(-1.68)	(0.74)	(1.59)	(2.71)
CMA	-0.38	0.87***	0.55***	0.93***	-0.01	0.40*	0.72***	0.73***
	(-1.54)	(4.57)	(3.17)	(3.80)	(-0.06)	(1.71)	(3.86)	(3.29)
WML	-0.12	-0.00	0.06	0.18*	-0.12	0.06	0.01	0.13
	(-1.16)	(-0.05)	(0.92)	(1.86)	(-1.32)	(0.78)	(0.18)	(1.39)
Adj. R2	0.69	0.70	0.70	0.23	0.72	0.72	0.70	0.25
Observations	132	132	132	132	132	132	132	132

Table 2. Carbon Intensity Sorted Stock Portfolios (cont'd)

Panel B: Equal-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Alpha	0.41**	-0.02	-0.11	-0.52***	0.35**	0.12	-0.18	-0.52***
	(2.05)	(-0.13)	(-0.69)	(-2.71)	(2.00)	(0.76)	(-1.15)	(-2.79)
MKT	0.84***	0.78***	0.84***	-0.00	0.85***	0.80***	0.81***	-0.04
	(17.05)	(17.36)	(20.39)	(-0.08)	(20.78)	(17.40)	(17.29)	(-0.78)
SMB	-0.72***	-0.37**	-0.26**	0.45***	-0.57***	-0.46***	-0.32***	0.25*
	(-3.94)	(-2.48)	(-2.37)	(3.02)	(-3.79)	(-2.93)	(-2.58)	(1.96)
HML	0.39**	-0.09	0.19*	-0.20	0.31**	0.22*	-0.05	-0.36**
	(2.18)	(-0.71)	(1.75)	(-1.15)	(2.08)	(1.81)	(-0.42)	(-2.10)
RMW	-0.59***	0.12	0.07	0.66***	-0.43**	-0.14	0.16	0.59***
	(-3.12)	(0.65)	(0.47)	(3.50)	(-2.47)	(-0.90)	(1.08)	(3.20)
CMA	-0.12	0.56***	0.39**	0.52*	0.02	0.19	0.62***	0.60**
	(-0.46)	(2.78)	(2.25)	(1.90)	(0.08)	(0.90)	(3.56)	(2.44)
WML	-0.14	-0.09	0.02	0.16*	-0.12	-0.04	-0.05	0.07
	(-1.32)	(-0.99)	(0.27)	(1.67)	(-1.32)	(-0.56)	(-0.76)	(0.93)
Adj. R2	0.69	0.68	0.75	0.19	0.75	0.71	0.72	0.24
Observations	132	132	132	132	132	132	132	132

Table 3. Carbon Intensity Sorted Dividend Strip Portfolio Expected Returns

This table reports annual expected returns on dividend strip portfolios sorted on carbon intensity. We retain December observations so that the one-year to four-year strips correspond to 12, 24, 36, and 48 months to maturity. At each December month-end, strips are assigned to tercile portfolios based on firm Scope 1 or Scope 2 intensity. Panel A reports value-weighted results using U.S.-dollar notional open interest, and Panel B reports equal-weighted results. Columns 1 through 4 correspond to one-year to four-year maturities. Newey-West t-statistics with a one-year lag are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels. The sample period is 2012 to 2022, with 11 annual observations per portfolio.

Panel A: Value-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	1-Year	**2-Year**	**3-Year**	**4-Year**	**1-Year**	**2-Year**	**3-Year**	**4-Year**
Group L	12.75***	11.93***	10.93***	9.03***	11.42***	11.59***	10.55***	7.24***
	(5.68)	(9.36)	(10.29)	(15.86)	(6.14)	(10.84)	(10.59)	(9.63)
Group 2	9.68***	9.74***	8.85***	9.76***	12.00***	10.27***	9.50***	10.75***
	(10.90)	(14.97)	(11.05)	(10.16)	(8.23)	(14.44)	(11.33)	(8.51)
Group H	9.21***	9.47***	9.04***	8.05***	8.56***	9.25***	9.16***	10.17***
	(8.13)	(11.56)	(11.02)	(7.88)	(7.80)	(9.46)	(14.56)	(6.86)
Group H-L	-3.54**	-2.46**	-1.90*	-0.98	-2.86**	-2.34**	-1.39*	2.93***
	(-2.26)	(-2.26)	(-1.80)	(-0.97)	(-2.17)	(-2.33)	(-1.95)	(2.59)
Panel B: Equal-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	1-Year	**2-Year**	**3-Year**	**4-Year**	**1-Year**	**2-Year**	**3-Year**	**4-Year**
Group L	18.98***	14.75***	12.92***	8.86***	17.19***	13.75***	11.69***	8.08***
	(8.52)	(11.03)	(10.04)	(12.90)	(9.06)	(12.71)	(13.16)	(6.67)
Group 2	13.66***	11.26***	9.62***	9.25***	16.05***	12.17***	11.02***	10.17***
	(10.04)	(9.34)	(14.24)	(8.43)	(10.97)	(11.37)	(11.85)	(9.08)
Group H	12.60***	10.06***	9.81***	7.62***	12.43***	10.68***	9.96***	8.98***
	(10.80)	(13.73)	(12.87)	(10.61)	(10.34)	(9.57)	(14.12)	(7.70)
Group H-L	-6.38***	-4.68***	-3.11**	-1.23**	-4.76***	-3.07***	-1.73***	0.90
	(-2.84)	(-3.18)	(-2.40)	(-2.05)	(-3.09)	(-5.94)	(-2.66)	(0.93)

Table 4. Carbon Intensity Sorted Dividend Strip Portfolio Expected CAPM Alpha

This table reports annual expected CAPM alphas on dividend strip portfolios sorted on carbon intensity. The portfolio construction follows Table 3. Panel A reports value-weighted results using U.S.-dollar notional open interest, and Panel B reports equal-weighted results. Columns 1 through 4 correspond to one-year to four-year maturities. Newey-West t-statistics with a one-year lag are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels. The sample period is 2012 to 2022, with 11 annual observations per portfolio.

Panel A: Value-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	1-Year	**2-Year**	**3-Year**	**4-Year**	**1-Year**	**2-Year**	**3-Year**	**4-Year**
Group L	8.95***	7.27***	5.52***	4.32***	7.72***	7.07***	5.26***	2.75***
	(3.83)	(5.75)	(6.02)	(6.51)	(3.87)	(6.41)	(6.31)	(4.54)
Group 2	9.83***	8.68***	6.76***	7.06***	11.81***	8.51***	6.60***	7.02***
	(10.03)	(14.09)	(9.21)	(7.97)	(7.77)	(10.65)	(8.18)	(8.03)
Group H	7.95***	7.18***	5.94***	4.06***	7.57***	7.34***	6.35***	6.37***
	(7.09)	(8.74)	(7.60)	(4.61)	(7.54)	(7.61)	(10.81)	(4.26)
Group H-L	-1.00	-0.09	0.43	-0.26	-0.15	0.27	1.10**	3.63***
	(-0.53)	(-0.09)	(0.48)	(-0.23)	(-0.10)	(0.25)	(2.04)	(2.80)
Panel B: Equal-Weighted								
	Scope 1 Intensity				**Scope 2 Intensity**			
	1-Year	**2-Year**	**3-Year**	**4-Year**	**1-Year**	**2-Year**	**3-Year**	**4-Year**
Group L	15.37***	10.85***	7.98***	4.41***	13.98***	10.05***	6.98***	3.90***
	(6.70)	(7.65)	(5.87)	(7.08)	(7.77)	(9.89)	(8.17)	(3.65)
Group 2	11.36***	8.87***	6.99***	6.36***	13.01***	9.01***	7.38***	6.22***
	(8.99)	(8.47)	(10.67)	(6.37)	(8.04)	(7.84)	(7.61)	(8.19)
Group H	9.98***	6.87***	6.40***	3.96***	10.31***	8.24***	7.28***	5.64***
	(9.53)	(9.48)	(9.86)	(6.65)	(11.04)	(9.09)	(11.07)	(4.41)
Group H-L	-5.39**	-3.98**	-1.58	-0.45	-3.67**	-1.81***	0.29	1.74
	(-2.16)	(-2.31)	(-1.14)	(-0.69)	(-2.37)	(-3.33)	(0.40)	(1.45)

Table 5. Carbon Intensity-Sorted Long-Term Asset Portfolio Results

This table reports long-term asset portfolio results sorted on carbon intensity. The left half reports portfolios formed on Scope 1 intensity, and the right half reports portfolios formed on Scope 2 intensity. At the beginning of each month, long-term assets are assigned to tercile portfolios based on stock emission-intensity measures observed at the end of the previous month. Panel A reports value-weighted results, where portfolio returns are weighted by U.S.-dollar stock market capitalization, and Panel B reports the corresponding equal-weighted results. In each panel, subsection .1 reports annual expected returns using December month-end observations only, as in Tables 3 and 4, whereas subsection .2 reports monthly realized return results using the full monthly sample as in Table 2. Realized return alphas and factor loadings are estimated from monthly regressions on the Fama-French six factors (Fama and French, 2018). For brevity, Panel B, subsection .2 reports only the FF6 alpha and omits the factor loadings. Newey-West t-statistics with one annual lag are reported for expected returns, and Newey-West t-statistics with three monthly lags are reported for realized return regressions. t-statistics are reported in parentheses below the point estimates. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. The sample period is 2012 to 2022.

Panel A: Value-Weighted Results								
	Scope 1 Intensity				**Scope 2 Intensity**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Panel A.1: Annual Expected Return								
Expected Return	2.4	8.14***	7.78***	5.38***	2.74	7.52***	8.49***	5.75***
	1.05	5.38	4.44	7.2	1.16	5.35	5.1	4.92
Obs.	11	11	11	11	11	11	11	11
Panel A.2: Monthly Realized Return								
Alpha	0.08	-0.18	0.09	0.02	0.28	-0.02	-0.21	-0.49
	(0.28)	(-0.73)	(0.40)	(0.05)	(0.96)	(-0.08)	(-0.83)	(-1.40)
MKT	1.08***	0.78***	0.87***	-0.21*	1.06***	0.86***	0.83***	-0.23**
	(10.41)	(11.54)	(14.81)	(-1.89)	(10.52)	(11.80)	(12.47)	(-2.20)
SMB	-0.75***	-0.58***	-0.71***	0.04	-0.57**	-0.76***	-0.68***	-0.10
	(-2.64)	(-2.69)	(-4.98)	(0.14)	(-2.15)	(-3.71)	(-3.95)	(-0.44)
HML	1.04***	-0.24	-0.05	-1.08***	0.82***	-0.12	0.06	-0.76**
	(3.67)	(-1.03)	(-0.23)	(-3.27)	(3.02)	(-0.47)	(0.26)	(-2.27)
RMW	-0.09	-0.04	-0.24	-0.16	-0.15	-0.40	0.22	0.37
	(-0.30)	(-0.14)	(-1.05)	(-0.50)	(-0.48)	(-1.39)	(0.70)	(0.95)
CMA	-0.70*	0.56*	0.96***	1.66***	-0.66*	0.72*	0.69**	1.35***
	(-1.78)	(1.71)	(3.50)	(3.62)	(-1.93)	(1.90)	(2.29)	(3.28)
WML	-0.20	-0.11	0.00	0.21	-0.09	-0.20	0.19	0.28
	(-1.22)	(-0.63)	(0.03)	(1.11)	(-0.58)	(-1.04)	(1.14)	(1.20)

Adj. R^2	0.65	0.53	0.62	0.29	0.66	0.54	0.50	0.25
Obs.	119	119	119	119	122	122	122	122

Table 5. Carbon Intensity-Sorted Long-Term Asset Portfolio Results (cont'd)

Panel B: Equal-Weighted Results								
	Scope 1 Intensity				**Scope 2 Intensity**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Panel B.1: Annual Expected Return								
Expected Return	0.74	7.79***	7.47***	6.73***	1.54	6.44***	7.91***	6.37***
	0.29	4.85	4	8.17	0.61	4.36	3.85	4.95
Obs.	11	11	11	11	11	11	11	11
Panel B.2: Monthly Realized Return								
Alpha	0.20	-0.09	0.21	0.00	0.54*	0.11	-0.16	-0.70
	(0.66)	(-0.29)	(0.79)	(0.01)	(1.79)	(0.41)	(-0.47)	(-1.63)
Obs.	119	119	119	119	122	122	122	122

Table 6. Carbon Emission Level-Sorted Portfolio Results

This table reports portfolio results sorted on carbon emission levels. The left half reports portfolios formed on Log Scope 1 emissions, and the right half reports portfolios formed on Log Scope 2 emissions. For stock and long-term asset portfolios, sorting variables are measured at the end of the previous month. For dividend strip portfolios, annual expected return and expected CAPM alpha are evaluated using December month-end observations only. Panel A reports value-weighted results. For stock and long-term asset portfolios, returns are weighted by U.S.-dollar stock market capitalization. For dividend strip portfolios, expected returns and expected CAPM alphas are weighted by U.S.-dollar notional open interest. Panel B reports the corresponding equal-weighted results. In the stock and long-term asset regression blocks, only alpha, adjusted R-squared, and observations are reported. In the dividend strip blocks, only the high-minus-low spread is reported across one-year to four-year maturities. Newey-West t-statistics are reported in parentheses below the point estimates. The sample period is 2012 to 2022.

Panel A: Value-Weighted Results								
	Log Scope 1 Emissions				**Log Scope 2 Emissions**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Panel A.1: Stock Monthly Realized Return Regression								
Alpha	0.41**	-0.04	-0.03	-0.45**	0.31*	-0.04	0.10	-0.21
	(2.11)	(-0.32)	(-0.22)	(-2.12)	(1.90)	(-0.25)	(0.67)	(-1.15)
Adj. R^2	0.69	0.72	0.71	0.18	0.74	0.70	0.72	0.14
Obs.	132	132	132	132	132	132	132	132
Panel A.2: Dividend Strip Expected Return (H-L)								
H-L	-3.14**	-2.06**	-1.84*	0.79	-1.59*	-0.36	-0.47	2.06
	(-2.07)	(-2.04)	(-1.85)	(0.38)	(-1.73)	(-0.42)	(-0.37)	(1.43)
Panel A.3: Dividend Strip Expected CAPM Alpha (H-L)								
H-L	-0.25	0.53	0.49	1.00	3.10***	2.94***	2.37**	2.70
	(-0.16)	(0.60)	(0.57)	(0.67)	(3.38)	(3.79)	(2.21)	(1.64)
Panel A.4: Long-Term Asset Annual Expected Return								
Expected Return	1.2	8.11***	8.0***	6.81***	2.55	6.4***	8.28***	5.73***
	(0.47)	(5.51)	(5.32)	(5.74)	(1.02)	(3.09)	(5.44)	(3.33)
Obs.	11	11	11	11	11	11	11	11
Panel A.5: Long-Term Asset Monthly Realized Return Regression								
Alpha	0.20	-0.35	0.01	-0.19	0.32	-0.38	-0.09	-0.41
	(0.66)	(-1.31)	(0.03)	(-0.48)	(0.88)	(-1.52)	(-0.31)	(-0.92)
Obs.	124	124	124	124	128	128	128	128

Table 6. Carbon Emission Level-Sorted Portfolio Results (cont'd)

Panel B: Equal-Weighted Results								
	Log Scope 1 Emissions				**Log Scope 2 Emissions**			
	L	**M**	**H**	**H-L**	**L**	**M**	**H**	**H-L**
Panel B.1: Stock Monthly Realized Return Regression								
Alpha	0.36*	-0.10	0.02	-0.33*	0.30*	-0.05	0.03	-0.27
	(1.86)	(-0.70)	(0.15)	(-1.84)	(1.71)	(-0.33)	(0.18)	(-1.59)
Adj. R^2	0.71	0.70	0.75	0.13	0.72	0.75	0.72	0.09
Obs.	132	132	132	132	132	132	132	132
Panel B.2: Dividend Strip Expected Return (H-L)								
H-L	-6.89***	-4.45***	-3.47**	-0.14	-4.19***	-2.08***	-0.70	0.97
	(-4.00)	(-3.49)	(-2.55)	(-0.07)	(-2.96)	(-4.05)	(-1.11)	(0.62)
Panel B.3: Dividend Strip Expected CAPM Alpha (H-L)								
H-L	-5.87***	-3.92***	-2.18	0.11	-2.58*	-1.60***	0.61	0.88
	(-2.98)	(-2.68)	(-1.54)	(0.07)	(-1.81)	(-2.72)	(0.90)	(0.52)
Panel B.4: Long-Term Asset Annual Expected Return								
Expected Return	0.01	7.72***	7.77***	7.76***	1.04	5.44***	8.09***	7.05***
	(0)	(5.07)	(4.6)	(6.79)	(0.4)	(2.67)	(4.29)	(4.6)
Obs.	11	11	11	11	11	11	11	11
Panel B.5: Long-Term Asset Monthly Realized Return Regression								
Alpha	0.30	-0.31	0.11	-0.19	0.40	-0.07	-0.10	-0.51
	(0.93)	(-1.06)	(0.35)	(-0.45)	(1.11)	(-0.25)	(-0.32)	(-1.11)
Obs.	124	124	124	124	128	128	128	128

The evolution of the TNFD: an institutional work perspective

Sisi Wu

Abstract

This study examines why and how actors from different fields collaborate to shape the sustainability reporting field. Through examining the evolution of the emergent Taskforce on Nature-related Financial Disclosures (TNFD), the paper attempts to reveal the relation dynamics during three interrelated phases: preparatory, design and adoption. Although the TNFD received great industrial adoption and government endorsements in a relatively short time, we have little understanding of its formation. The research puts emphasis on addressing how core tensions were raised and strategies employed by the TNFD at each stage to mobilise a wide range of supporters as well as accommodate competing interest groups who seek to promote their own positions in specific directions (Gross and Zilber, 2020). More specifically, it mobilises Lawrence and Suddaby's (2006) notion of institutional work typology to study how the TNFD enters and reshapes the nature-related risk reporting landscape through strategies such as advocacy, defining and mythologising. This study employs an in-depth qualitative case study, relying primarily on a wide range of documentary sources, including the TNFD reports, discussion papers, comment letters, press releases, government announcements and transcripts of public webinars.

While most research has focused on climate reporting, this research enriches knowledge of how the nature-related reporting framework was produced. As the first in-depth case study of the TNFD's formation, it extends understanding of how a voluntary reporting recommendation gains legitimacy and the behind-the-scenes interplay of diverse actors. This study also highlights the strategic collaboration between diverse actors, especially NGOs and private financial institutions, to shape the sustainability reporting field. Lastly, this paper shows how the TNFD ambitiously intended to provide more precise clarification but then introduced ambiguity and flexibility into its framework.

1. Introduction

“There is no prosperity without nature.”

-Marcos Neto, UN Assistant Secretary-General

The resilience of nature determines the future of all creatures on the earth. As a crucial part of the ecosystems, natural resources not only underpin every aspect of human health and living quality, but they also contribute to society and economic development (UNDP, 2021). However, due to the influence of human activities, nature is fading rapidly at an unexpected rate (WWF, 2024). For example, the global forest cover in the expansion of agricultural production has declined by 26 per cent over the past decade, devastating the industries that rely on it (UNEP, 2021). Back in the last century, the Brundtland Commission Report (1987) had already demonstrated an acute awareness of the interconnected challenges posed by sustainability and future development, issuing an urgent appeal for coordinated global efforts across sectors. This also led to the immense scrutiny of business models prioritising short-term economic growth (Laine et al., 2021). In response to growing public scepticism and building reputation (Cooper and Owen, 2007), organisations have begun to provide sustainability information, which usually discloses operations, strategies and investments related to sustainability to broad stakeholders (Laine et al., 2021). Consequently, a variety of frameworks and guidance sprung up, such as the Global Reporting Initiative (GRI) Standards, European Sustainability Reporting Standards (ESRS) and the International Sustainability Standard Board (ISSB) Standards. While the majority of frameworks claim on their websites that their recommendations could contribute to reporting transparency, consistency and comparability, the divergence in priorities, target users and measuring metrics bring a fragmental and complex picture in the sustainable reports field (Rowbottom, 2022; Baudot, 2014), which has even been described as the “alphabet soup” (Pelger et al., 2024). Although concerns about climate risk have been burgeoning in recent years, the endeavours on nature-related issues are still scant. For nature and biodiversity[1], it is much more complex and multifaceted. Unlike climate change, which can be broadly measured through greenhouse gas emissions, nature-related issues are highly context-specific, varying by geography and ecosystem (Global Canopy, 2020). Accordingly, there is a lack of research on nature-related reporting disclosures. Despite this new terrain might not be there for accounting professionals to consider today, it will bring countless opportunities and challenges soon (Laine, 2024).

In this context, the Taskforce on Nature-related Financial Disclosures (TNFD) was established in 2021 to create a set of consistent disclosure frameworks that allow businesses to incorporate nature-related risks and opportunities into their strategic

[1] Nature and biodiversity: These two words are interchangeable in this paper.

planning and ensure better information to capital providers (TNFD, 2025). It is the first initiative that focuses particularly on nature issues. Thus, there is a lack of knowledge regarding the development of nature-related risk reporting. In addition, while frameworks such as the TCFD are sector-based (O'Dwyer and Unerman, 2020), the TNFD's development involves a different set of stakeholders under its claimed market-led approach, including non-governmental organisations (NGOs), corporates, financial institutions, governments, accounting service providers, data providers and local communities. This makes the TNFD disclosure generation process noisier and more complex. Hence, this research studies the emergence of the TNFD and attempts to reveal the relation dynamics and conflicts during this process. Until 2024, the TNFD has received a great number of adoptions from industry, with over 500 corporates and financial institutions from 49 jurisdictions and across 62 sectors adopting the TNFD recommendations, representing over 6.5 trillion dollars in market value until this year (TNFD, 2024b). Rather than creating a brand-new set of guidelines, the TNFD was built upon other existing frameworks, especially by emulating the Taskforce on Climate-related Financial Disclosures (TCFD). However, building on other mature frameworks does not mean there is less confrontation in the creation of the TNFD. On the contrary, actors with competing interests who are temporarily bound will seek to promote their own positions and configure the fields in specific directions (Gross and Zilber, 2020).

By detailedly examining archival data from diverse sources, it investigates how and why actors with divergent interests, such as NGOs and private financial institutions, negotiate to shape the nature-related risk reporting field. Drawing on the evolution of the TNFD, the paper attempts to reveal the relation dynamics during three interrelated phases: preparatory, design and adoption. It puts emphasis on addressing how core tensions were raised and strategies used by the TNFD at each stage to mobilise a wide range of supporters. The story of the TNFD's formation extends the understanding of how a voluntary reporting recommendation gains legitimacy and the behind-the-scenes interplay of actors. It also explores how the TNFD oscillated between clarification and ambiguity, thereby ensuring broad participation in the burgeoning sustainability reporting field.

The research employs the notion of institutional work typology that was initially brought by Lawrence and Suddaby (2006), with a particular focus on creating and maintaining institutions. The institutional work theory focuses on the intentional actions of individuals and organisations in shaping organisations. This aligns with the TNFD's evolution, where a diverse set of actors actively engage in various forms of institutional work to legitimise, contest or affect the reporting framework. It enables a more nuanced understanding of actions underlying institutional emergence and development. Although the TNFD received such great support and endorsement in a relatively short time, we know little about the story of the TNFD's emergence, especially the interplay of conflicting actors within the framework-setting process of the TNFD formation. By using an institutional work lens, it advances existing literature by analysing the interplay and dynamics between these two institutional processes. At the same time, the research integrates multiple forms of work across these stages and renames them into three functional categories that explain how actors manage tensions and secure legitimacy.

This paper seeks to contribute to the existing literature through three dimensions. Firstly, it expands the understanding of the emergence and evolution of the TNFD. While most research has focused on climate reporting, this research enriches knowledge of how nature-related reporting framework was produced. Although some prior accounting literature has addressed the formation process of initiatives within the sustainable reporting field, the TNFD evolution itself is distinct. Different from extant reporting frameworks, the TNFD confronts the inherent complexity and geographic specificity. Such characters hinder standardisation, rendering the development of globally recognisable nature disclosure particularly difficult. Moreover, the TNFD's formation process is distinguished by a deliberate attempt to include a wide range of actors with different interests and values, challenging the traditional standard-setting models. As common grounds or collective benefits are required to overcome the consensus dilemma (Gray and Purdy, 2018), hence, this research investigates how diverse actors especially NGOs and private financial institutions collaborated to shape the sustainability reporting field. It puts emphasis on addressing how core tensions were raised and strategies employed by the TNFD to mobilise a wide range of supporters and accommodate divergent interests. It aims to provide valuable insights into the relation dynamics during the framework-setting process and how a voluntary framework gains legitimacy and evolves. Secondly, this paper extends Lawrence and Suddaby's (2006) institutional work taxonomy. While Lawrence and Suddaby (2006) admit this is a preliminary classification, this paper further develops an understanding of the interaction between different forms of institutional work (Lawrence et al., 2013; Currie et al., 2012; Zilber, 2002). By examining the TNFD's iterative evolving process under its open innovation approach, this research extends previous arguments by further examining how institutional creation and maintenance can occur simultaneously and deeply intertwined rather than as sequential or separate stages (Canning and O'Dwyer, 2016; Hayne and Free, 2014; Jabbour et al., 2025; Farooq and Villiers, 2018). To capture these insights, this paper introduces a refined model that structures institutional work into three interrelated functional categories that are summarised from the case narratives: legitimacy building, incremental permeation and coalition creation. This model moves beyond a static classification of institutional work to reveal how institutions are not only built but also actively sustained during their formation by observing how these categories integrate multiple forms of work. It also offers a novel framework for studying the institutionalisation process in the framework-setting process of a voluntary reporting initiative. Lastly, it explores how a delicate balance between ambiguity and clarification has been determined by the TNFD. Prior research suggests that the emergence of a new institution could benefit from flexibility and ambiguity (O'Sullivan and O'Dwyer, 2015; Baudot, 2014; Fernández Chulián et al., 2024), allowing diverse actors to reach a consensus on a broad and flexible level.

However, institutions must provide clarity (Wijen, 2014; O'Dwyer et al., 2024) at the same time to support adoption and sustain legitimacy. In the case of the TNFD, it shows an opposite process. The TNFD ambitiously and boldly intended to provide more precise clarification such as the narrowed focus of financial-oriented materiality and the staged appliance approach. Nevertheless, clarity may not be the endpoint of a linear process, but rather a strategy used at initial days to attract enrolment. As the TNFD framework evolved, its initial claims to clarity gradually gave way to a more ambiguous and flexible design logic. This move reflects a strategic shift and challenges the previous impression of the relationship between clarity and ambiguity.

2. Literature Review

The global architecture of accounting standard-setting has undergone a fundamental transformation over the past two decades. There is an inexorable rise of private standard and framework setters which immensely reshaped the infrastructure of accounting governance (Chua and Taylor, 2008); for example, prominent private organisations such as the International Accounting Standards Board (IASB) and the Financial Accounting Standards Board (FASB) play a dominant role by developing pre-formulated frameworks which governments may adopt later and thus affect regulatory outcomes (Chiapello and Medjad, 2009). However, unlike national standard setters that automatically obtain prestige and government endorsements (Pelger and Spieb, 2016; Richardson and Eberlein, 2010), these private institutions usually cumulative their legitimacy and secure support by increasing user participation and engaging in inclusive consultation processes (Durocher et al., 2007; Allini et al., 2018). This shift has made the standard-setting process increasingly complex and fragmented where competing interests, values and political purposes converge in the space of accounting. As a result, the accounting system is no longer just about the standard makers (Slager et al., 2012) but is more likely to be a carrier where interests are negotiated and political concerns are expressed (Burchell et al., 1980). The intersection of accounting and sustainability governance makes this problem more protrude. Therefore, due to the heterogeneous interests, it is extremely challenging to embed diverse values and purposes into a single set of accounting frameworks (Arjaliès et al., 2023; Bellucci et al., 2019) and achieve consensus (Brunsson et al., 2012). To understand this dynamic, it is necessary to have a closer examination of how voluntary frameworks come into being and evolve is necessary.

Building on the expanding literature examining the development of sustainability reporting, recent research has begun to explore how sustainable standards and frameworks evolve through contested and negotiated processes. For example, Clune and O'Dwyer (2020) illuminate how an advocacy movement formed to introduce environmental and social accountability mechanisms in the Dutch investment field. They identified meaningful collaboration and temporary agreements between actors under a series of educational work, ranking work and soft advocacy work that allows opposing value frames to co-exist peacefully. Similarly, Humphrey et al. (2016) capture how the International Integrated Reporting Council seeks to gain a wide range of supporters by ratifying reciprocal agreements with a large number of paramount institutions in the reporting field. They also indicate shifts from distinguishing its work from other forms of disclosing recommendations to gradually downplaying its distinction to foster collaboration. Consequently, this shifting of strategic emphasis led to its demise (O'Dwyer et al., 2024). Furthermore, Maechler and Graz (2019) put forward that the standardisation of natural capital accounting is not only a technical process but also involves political contests, especially power relations. Although the process of accounting standard-setting by private bodies has received some attention in recent years, as claimed by O'Dwyer et al. (2024), the practice of handing over the power to these institutions remains far from settled, especially under the situation that the TNFD is a voluntary initiative but positioned itself as multi-stakeholder oriented. This challenges the predominately investor focus observed in prior studies and enables a deeper understanding of how claims to inclusiveness are constructed and implemented in practice. The TNFD thus provides a valuable case for interrogating how power is contested and negotiated in accounting standard-setting, particularly about whose voices are emphasised while the others are ignored, and how legitimacy is produced and developed. Besides, there is a gap in understanding the issues that emerge when transferring the climate risk reporting framework into the more complex and fragmented biodiversity field. The burgeoning nature-specific framework is in reality more complicated. The TNFD's evolution provides a revealing site through which to examine how the familiar disclosure framework is reinterpreted in response to ecological uncertainty and stakeholder contestation, gaining insight into what particular considerations need to be incorporated into the comprehensive understanding of the sustainability disclosure area. This research also further examines the limited field-level interplay (O'Dwyer, 2023; O'Dwyer et al., 2024; Zietsma et al., 2017) within the TNFD context and exploring the institutional tensions that shape the development of biodiversity disclosure frameworks. Besides, due to the nature of existing initiatives, they rarely dig into the internal interactions between NGOs and private financial institutions, such as the motivations driving their cooperation and the significance of these interactions. In doing so, the paper sheds light on how institutional tensions are negotiated in the early stages of sustainability framework-setting, especially with biodiversity and ecosystem valuation.

3. Theoretical Framework

To explore the conflicting nature of the TNFD formation process and how actors with divergent purposes could work together towards the same goal, this study draws on the notion of Lawrence and Suddaby's (2006) institutional work typology. It provides a structured way to analyse how institutions are created, maintained and disrupted. By adopting an institutional work lens, it allows an examination of how the TNFD enters and reshapes the nature-related risk reporting landscape through strategies such as advocacy, defining and mythologising. Given the disorder that arises during the emergence of the TNFD, institutional work forms facilitate to capture of the complex actor actions and their dynamic interrelationship (Battilana et al., 2009).

3.*1 The notion of institutional work*

The notion of an institution is situated at the centre of organisational research and has been discussed extensively. A common definition was put forward by Scott (1995), referring to institutions as complex structures and frameworks constituted from ‘regulative, normative and cultural-cognitive factors’ that guide people’s behaviours. Over the past few decades, researchers’ perspectives on the functioning of institutions have undergone significant evolution. While the old institutionalism primarily stresses the power and formal structures (Selznick, 1949; Clack, 1960; Meyer and Rowan, 1977), the new institutionalism shifts the focus and introduces the cultural, symbolic and cognitive dimensions in shaping institutional dynamics (DiMaggio and Powell, 1983; Greenwood and Hinings, 1996). This shift enables closer investigation into the daily practices and informal negotiations (Scott, 1995) that shape the institutions. However, the new institutionalism has been criticised for its excessive focus on the influences of institutions on human actions, while ignoring the reciprocal impacts of human agency on institutional structure (Empson et al., 2013).

Lawrence and Suddaby (2006) conceived the institutional work as ‘the purposive action of individuals and organisations aimed at creating, maintaining and disrupting institutions’. This offers a detailed account of the day-to-day practices and strategies employed by actors to influence institutions. Echoing the new institutionalism, their viewpoint also stresses the role of human agency which actively constructs, adapts and may even disrupt institutional arrangements. Thus, it highlights the interrelations between actors’ actions and institutional dynamics (Battilana et al., 2009). Lawrence et al. (2011) further claim that the institutional work examines how and why actors interpret, transform, edit and remodelling institutions and how these actions lead to certain consequences. The key matter behind institutional work is to understand how and why actors carry out work in such a way and the consequences of their actions (Zietsma and Lawrence, 2009). Therefore, by utilising the concept of institutional work, researchers can analyse the institutional process through a fluid and uncertain lens (Lounsbury, 2008). However, it is worth noting that institutional work is not always associated with successful attempts to change the institutions. The institutional work describes purposive action aimed at affecting the institution, no matter whether it succeeded, failed or has no effect (Lawrence et al., 2009). Thereout, focus should be put on the work itself rather than accomplishments (Lawrence et al., 2013; Empson et al., 2013). Moreover, the institutional work also highlights the reflexivity of both individual and collective actions (Lawrence et al., 2013; Canning and O’Dwyer, 2013), which highlights actors actively shaping the institutional structures and adapting their strategies to respond to challenges.

3.2 Institutional work taxonomy

Lawrence and Suddaby (2006) initially divided the institutional work into three categories: creating, maintaining and disrupting institutions, with diverse forms of work within each category (See Table 1). It attempts to capture the complex actor actions and their dynamic interrelationship (Battilana et al., 2009). The first category creating institutions includes three sets of institutional work: overtly political work (advocacy, defining and vesting), reconfiguring actors' belief systems (constructing identities, changing normative associations and constructing normative networks) and altering meaning systems (mimicry, theorising and educating) (Lawrence and Suddaby, 2006). The ability to create an institution relies on establishing rules and employing rewards and punishments to ensure that those rules are enforced (Lawrence and Suddaby, 2006). However, the key to such ability always links to the resources actors hold (Pfeffer and Salancik, 1978) and specialised identity (Lawrence and Suddaby, 2006). This means actors with less power usually try to build institutions by relating new institutions to existing practices (Lawrence and Suddaby, 2006). The second category, maintaining institutions, received much less attention compared to the focus on building new institutions (Scott, 1995). Maintenance work is not just reproducing taken-for-granted rules (Jepperson, 1991; Hwang and Colyvas, 2010). Nevertheless, many scholars claim the maintaining stage is not static but relatively fluid. It requires sustaining, repairing and reconstructing institutional mechanisms that can be achieved through either reinforcing the rule systems (Suddaby and Greenwood, 2005; Zilber, 2009) or preserving and legitimating institutional authority (Lawrence and Suddaby, 2006). Micelotta et al. (2013) further developed the maintaining work and their study shows how precarious institutional arrangements can be repaired by the efforts of incumbent actors, reversely transformed to rebuild the status quo. The last category, disrupting institutional work aims at breaking or impairing institutions when actors' interests are not served by existing institutions (Power and DiMaggio, 1991). Although some researchers only mention actors' disruptive efforts when emphasising the process of institutional creation, the disruption work itself is distinct and extends beyond merely facilitating the emergence of new institutions (Oliver, 1992). This kind of work usually involves lowering the costs of non-compliance and violation, thereby separating rewards and sanctions from current institutional arrangements (Lawrence and Suddaby, 2006). In addition, the use of diverse forms of work in disrupting institutions depends on the nature of different actors (Lawrence and Suddaby, 2006). For example, while the states and governments hold the power to directly detach the rewards or sanctions from the institutions, less powerful actors tend to undermine the institutions by gradually undermining the belief systems. However, this research will not focus on disrupting institutions. Echoes with earlier arguments, disruptive institutional work can be considered as purposive work that occurs when actors intend to break down the existing status quo (Zietsma and Lawrence, 2010). The TNFD was trying to develop its framework upon other existing initiatives, therefore, it adopted a relatively collaborative and imitative strategy, without a conspicuous intention to break other mature initiatives. Moreover, the sustainable reporting field is marked by fragmentation and diversification, lacking a single dominant actor (Zietsma et al., 2017). In such a setting, disruptive action aimed at displacing a prevailing framework setter would be either infeasible or counterproductive. In summary, this paper concentrates on actions that aim at creating and maintaining institutions and their relations.

Creating Institutions	
Advocacy	The mobilisation of political and regulatory support through direct and deliberate techniques of social suasion
Defining	The construction of rule systems that confer status or identity, define boundaries of membership or create status hierarchies within a field
Vesting	The creation of rule structures that confer property rights
Constructing identities	Defining the relationship between an actor and the field in which that actor operates
Changing normative associations	Re-making the connections between sets of practices and the moral and cultural foundations for those practices
Constructing normative networks	Constructing of interorganisational connections through which practices become normatively sanctioned and which form the relevant peer group with respect to compliance, monitoring and evaluation
Mimicry	Associating new practices with existing sets of taken-for-granted practices, technologies and rules in order to ease adoption
Theorizing	The development and specification of abstract categories and the elaboration of chains of cause and effect
Educating	The educating of actors in skills and knowledge necessary to support the new institution
Maintaining Institutions	
Enabling work	The creation of rules that facilitate, supplement and support institutions, such as the creation of authorizing agents or diverting resources
Policing	Ensuring compliance through enforcement, auditing and monitoring
Deterring	Establishing coercive barriers to institutional change
Valorising and demonizing	Providing for public consumption positive and negative examples that illustrates the normative foundations of an institution
Mythologizing	Preserving the normative underpinnings of an institution by creating and sustaining myths regarding its history
Embedding and routinizing	Actively infusing the normative foundations of an institution into the participants' day to day routines and organisational practices
Disrupting Institutions	
Disconnecting sanctions	Working through state apparatus to disconnect rewards and sanctions from some set of practices, technologies or rules
Disassociating moral foundations	Disassociating the practice, rule or technology from its moral foundation as appropriate within a specific cultural context
Undermining assumptions and beliefs	Decreasing the perceived risks of innovation and differentiation by undermining core assumptions and beliefs

Table 1:Lawrence and Suddaby's (2006) Institutional Work taxonomy

The above institutional work taxology provides research with a consistent and clear map of the institutional work life cycle. However, as criticised by some scholars, considering the complexity and mess in reality (Cooper et al., 1996), the boundaries between these categories could be blurred and their relations are highly dynamic and fluid. For example, Perkmann and Spicer (2008) perceive that there is little clarification about how different forms of institutional work interact. According to Hayne and Free (2014), such categories are not precise, and this process could be unexpected and dynamic. Meanwhile, Canning and O'Dwyer (2016) illuminate that such broad classification 'lacks the nuance necessary' and fails to capture the inherent complexity which motivates actors to incite change. After embedding the abstract theory into empirical application, Empson et al. (2013) reveal that the oversimplified theory will lead to muddles and misunderstandings as it limits the notion of singular action by a single subject that intends to influence an institution. Some papers point out that creating and disrupting institutions could occur simultaneously when actors are pursuing changes (Jabbour et al., 2025; Farooq and Villiers, 2020). In addition, different forms of institutional work are not isolated from each other, and their application is not subject to the broad classification. Farooq and Villiers's (2018) paper points out that social actors may adopt a single form of institutional work at different stages. Moreover, the pursuit of greater institutional change or the presence of stronger resistance will increase the likelihood of multiple forms of institutional work occurring at the same time. Hayne and Free (2014) also perceive that different forms of institutional work can transcend categorical boundaries and interact with each other. There is also evidence indicating that the employment of one institutional work could support and reinforce another one (Micelotta et al., 2013).

3.3 Mobilising the creating and maintaining stages

In general, Lawrence and Suddaby's (2006) institutional work theory offers an approach to capture, organise, classify and explain the complex actors' actions and their implications on the institutions (Jabbour et al., 2025). However, the framework setting process is much more complicated, dynamic and nuanced, especially within the situations of diverse actor engagements. Therefore, the broad categories may need to be further adjusted to accommodate new circumstances. More research is required to explore the institutional dynamics that examine various forms of work within the same context (Lawrence et al., 2013), such as how different forms of institutional work could reinforce each other and the reasons for choosing specific forms over the others (Canning and O'Dwyer, 2016). Higgins et al. (2015) call for a refinement of the institutional theory to better understand sustainability reporting behaviours. As this research aimed at investigating the dynamics of the TNFD's emergence through observing the tensions and negotiations among various parties, mobilising the institutional theory provides a valuable lens for understanding the motivations, actions and strategies employed by different participants during this process. This research emphasises the creating and maintaining stages, and explores how these two stages intertwine. Before its official release, the TNFD recommendations underwent four draft stages. Following each round of feedback, the recommendations were revised and a new version was issued, with the process continuing iteratively. The ongoing beta stages not only continuously build new scaffolds, but also maintain and improve the outcome of previous versions. This cyclical approach blurred the boundaries between the creation and maintenance stages, resulting in more complex interactions among various forms of institutional work. Hence, this paper delves into the underdeveloped maintenance stage and illustrates its relations with the creating stage.

4. Research Method

4.1 Data collection

Given this research aims to deepen understanding of the TNFD's formation and evolution through an institutional theory perspective, an in-depth single qualitative case study is employed. Qualitative research focuses on understanding social actors' meanings and phenomena in particular environments (Denzin and Lincoln, 1994). Instead of attempting to generalise some common sense from the study, qualitative research delves into the uniqueness of the research phenomena, enabling a deeper and richer understanding of context-specific information (Dyer and Wilkins, 1991; Van Burg et al., 2020). Such research method supports addressing the "how" and "why" questions, which is consistent with the research objectives: How do core tensions arise during the TNFD framework setting process and how these contradictions are managed?

A wide range of documents were the primary data sources of this study. All data is publicly available and collected from diverse platforms, including formal TNFD reports published on its official websites such as the final Recommendations of the TNFD, Publication Archival of the drafted frameworks in all four testing stages, Sector guidance that especially for the financial industry, press release, Discussion papers that exploring key challenges, possibilities and improvements, meeting transcripts and educational webinar records. Emphasis will be placed on ensuring it aligns with the specific focus of the dynamics of the TNFD. In addition, the development timelines of previous reporting initiatives such as the GRI and IR will also be looked through to gain some background information and understand the uniqueness of the TNFD's creation. The interactions between the TNFD and other actors will also be captured through various documents, such as Standards alignment and mapping between the TNFD and other competing frameworks, Public comments letters and TNFD's response to these comments. This is continually updated based on new information published. To achieve the objective of comprehensively understanding the tensions and dynamics inherent in the TNFD recommendation-creating process, archival data from the websites of other significant actors will also be meticulously examined. For example, the pilot testing report by the United Nations Environment Programme Financial Initiative (UNEP FI), funding information from the Word Wide Fund for Nature and the sector pilot testing case studies led by Global Canopy. Lastly, engagements by professional accounting bodies including KPMG Sustainability, PwC, Institute of Charted Accountants in England and Wales (ICAEW), Association of Charted Certified Accountants (ACCA) and commentary from financial media such as Financial Times and Bloomberg News will also be analysed to capture diverse perspectives from industry experts, journalists, and commentators. The broad scope ensures the inclusion of viewpoints from key players who either greatly participate in creating the TNFD framework or have technical knowledge. The main evidence used is summarised in Table 2.

TNFD's published resources (examples):

- "Nature in scope" (26 pages)
- The TNFD framework beta v0.1 to v0.4 (179 pages in total)
- "Taskforce on Nature-related Financial Disclosures (TNFD) Recommendations" (154 pages)
- "Getting started with adoption of the TNFD Recommendations" (22 pages)
- "Additional guidance for financial institutions" (37 pages)
- "Discussion paper on nature transition plans" (106 pages)
- 27 published comment letters (151 pages)

Reports (examples):

- WWF, "Public Development Banks and Biodiversity" (37 pages)
- WWF, "The Nature of Risk" (42 pages)
- UNEP FI, "Unboxing Nature-related Risks: Insights from the UNEP FI-led TNFD Piloting Programme" (78 pages)
- UNEP FI, "Accountability for Nature: Comparison of Nature-Related Assessment and Disclosure Frameworks and Standards" (88 pages)
- The Biodiversity Consultancy, "TNFD Financial markets Readiness Assessment" (44 pages)
- WWF & AXA, "Into the Wild: Integrating nature into investment strategies" (41 pages)
- UNEF FI, "UNEP FI Pilots in Support of TNFD" (32 pages)
- KPMG, "The move to mandatory reporting: Survey of Sustainability Reporting 2024" (87 pages)
- UK Government, "Business case summary of the TNFD" (64 pages)

Conferences/Webinars (Examples):

- Green Finance Institute, "Panel discussion from 3 UK companies on their progress"
- Green Finance Institute, "Getting Started and Early Adoption"

- TNFD, Global launch event
- United Nations, “Biodiversity: Facing an existential crisis”

News/Commentary (Examples):

- Financial Times, “David Craig: Our economic system is completely dependent on nature”
- ICAEW, “Adopting the TNFD framework needn’t be a challenge”

Table 2: Some key data used in this research

Compared to other qualitative research methods, documentary analysis is efficient and cost-saving (Bowen, 2009). This method provides access to comprehensive coverage of materials, including reports, consultation papers, and news reference records, which are often easily retrieved online from official websites and do not require formal authorisation (Morgan, 2022). This makes it particularly advantageous in research contexts like the TNFD, where extensive publicly available documents exist to support the study of institutional processes and actor dynamics. In addition, one of the strengths of using officially published documents lies in their reliability and unaffected data (Merriam and Tisdell, 2016). These materials are usually subjected to a rigorous review process, which ensures a high degree of accuracy and minimises errors. Additionally, documentary records allow researchers to trace the development of ideas and frameworks over time and provide valuable insights into how priorities, purposes, and strategies have shifted throughout the TNFD’s formation and development. However, the documentary analysis is by no means free of issues. For example, publicly available documents may not always reveal the full spectrum of interaction, conflicts and negotiations experienced in the decision-making process (Blackstone, 2019). It is challenging to uncover deeper conflicts; therefore, this paper aims to maximise the use of publicly available materials to achieve the research purposes. To mitigate such risks and the limitations inherent in relying on a single data source, this research uses triangulation as a robust validation strategy by combining data from multiple sources for cross-verification (Flick, 2004). For example, while the TNFD reports claim broad stakeholder consensus, comment letters from NGOs might uncover significant unsolved concerns regarding meaningful engagements that are masked in official publications.

4.2 Data Analysis

The data analysis process was conducted concurrently with the ongoing data collection. It adopts a modified form of content analysis (Shapiro and Matson, 2008; Humphrey et al, 2016) to examine the archival data that was introduced earlier. This method involves a careful investigation of both manifest (literal) and latent (deep structural) meaning with the data (Shapiro and Matson, 2008; Suddaby et al., 2007), offering initial insights into the key events surrounding the establishment of the TNFD, the primary actors participated and main interaction ways between them. The analysis proceeded in three steps. Firstly, all documents were read line-by-line to construct an overall map of the storyline, while a detailed timeline of major events and activities was simultaneously developed (See Figure 1). The whole process was organised into three interrelated stages arranged in chronological order: the Preparatory Phase, which outlines how the idea for creating a task force around nature-related reporting was generated; the Design Phase, involving four beta versions leading up to the official publishing; and the Adoption Phase, which focuses on promoting its final recommendations. Following a comprehensive descriptive summary of key findings, a more interpretive narrative (O'Dwyer, 2004) that supports answering the research questions was carried out. Thus, a more narrowed focus was being put on examining the core tensions, negotiations and strategies that emerged. Relevant documents were further selected and organised using the software NVivo 12. It employed a combination of inductive (data-driven) and deductive (theory-driven) coding (see Fereday and Muir-Cochrane, 2006). Documents were re-read several times in order to capture codes that reflect the dynamics. Additionally, NVivo's text search function was utilised to identify interactions and negotiations around key terms such as "materiality", "ambiguity", "participation" and "feedback". The data analysis process still involves data collection. As the TNFD continues to evolve, the data analysis process remains iterative, with new evidence incorporated over time. In parallel, Lawrence and Suddaby's (2006) institutional work taxonomy and other subsequent literature were comprehended and mobilised in order to understand how actors could shape the institutions by diverse actions. To facilitate this, I pre-established a set of categories corresponding to different forms of institutional work and then clustered the nodes extracted from the documents within these categories. It requires an ongoing iterative dialogue between data and theory, whereby emerging insights could develop theories through problem-solving and utilise the existing frameworks (Alvesson and Karreman, 2007) and, conversely, theoretical ideas support data interpretation. In the last stage, these dispersive categories converged into three aggregated dimensions (Gioia et al., 2012), namely legitimacy building, incremental permeation and coalition creation. The details will be illuminated in the following sections.

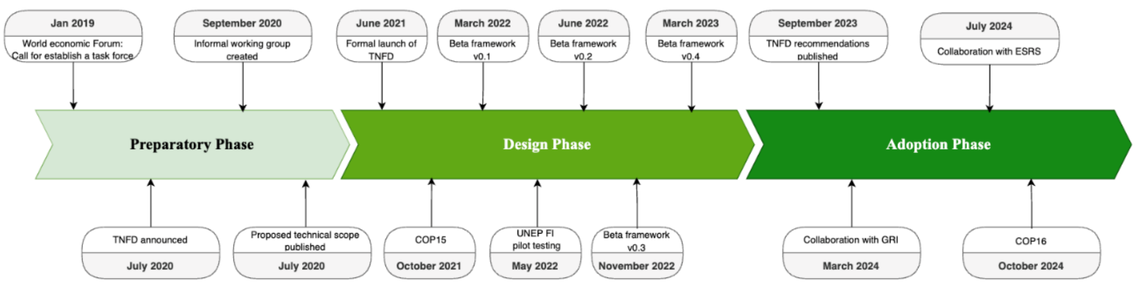


Figure 1: The timeline of the TNFD evolution (created by author)

5. Case Narrative

The case narrative section examines the formation of the TNFD through the lens of institutional work in three chronological phases: preparatory, design and adoption. Far from a neutral technical exercise, the TNFD's creation unfolded as an open, contested process that is buzzing with actors seeking to shape the framework. This paper highlights how different forms of institutional work co-evolve through iterative consultation, working groups, pilot testing and negotiation. It also reveals how core tensions arise at the initial stage, efforts made to facilitate conflicts and the dynamic relationship between diverse actors.

5.1 The preparatory phase

5.1.1 Impulse for creating the taskforce: stressing the urgency of nature issues

Although there was mounting scientific evidence of ecosystem collapse, biodiversity loss initially struggled to capture the attention of the financial community. With the continuous advancement of climate disclosures as a pioneer, other sustainability subfields such as the nature-related risk reporting field eventually started to sprout. In 2017, 50 conservation scientists proposed a more active response to escalating biodiversity loss and brought up the idea of establishing a new 'Global Deal for Nature' as a companion to the Paris Climate Agreement (WWF, 2018). Subsequently, two global emerging sustainability policies, the Convention on Biological Diversity and the Sustainable Development Goals, gained traction and positioned solving nature issues as a long-term target. The non-governmental organisation World Wildlife Fund (WWF[2]) immediately caught the tide and issued its *Living Planet Report 2018,* which called for a most ambitious global agreement that gathers a consortium of elites to create a new deal for nature and people (WWF, 2018). This proposition laid the very initial prototype for what would later evolve into the TNFD. Therewith, its idea was further prompted in January 2019 at the World Economic Forum's Davos[3] meeting (UNEP FI, 2021):

"We are losing nature too fast. We need a global compass for nature to be agreed in 2020 by governments and business, as we did for climate in Paris (WWF, 2020)."

This meeting was the first time the non-governmental organisation had engaged in such international forums that attempted to influence the world's powerful organisations and governments. The Davos 2019 thereafter was seen as the starting

[2] The WWF is the world's leading conservation organisation with net assets of 623 million dollars and working with a wide range of organisations to support around 3000 conservative and environmental projects.

[3] The World Economic Forum is the International Organization for Public-Private Cooperation. It provides a global, impartial and not-for-profit platform for meaningful connection between stakeholders to establish trust and build initiatives for cooperation and progress (World Economic Forum, 2024).

point for nature and biodiversity loss featured prominently in the overall programme. The potential task force, claimed by WWF (2019), aimed to raise awareness of nature through this highly influential audience. By stressing the urgency of nature-related issues during the event, the WWF also asked businesses, financial institutions, and governments to work together and achieve such ambitious goals:

“We need governments, businesses, financial institutions, civil society and everyone to commit to halt and start reversing the loss of nature. At Davos, we will be asking business and political leaders to come together to help achieve this aim. Together, we can and will save our planet (WWF, 2019).”

Soon afterwards, WWF France co-published a report with AXA, a multinational insurance corporation based in France. Their collaboration may appear surprising due to their seemingly irreconcilable interests. While WWF tends to stand for the public and environment as an influential NGO, AXA is embedded with a market logic that is oriented towards profitability. However, WWF realised that reversing nature loss requires a collective global effort that extends beyond environmental organisations. The material and symbolic support it gained from multinational financial institutions will give it further authority on an international stage (Ronzani and Tweedie, 2024). As it explained:

“We need a global action to bend the curve of nature loss, protect our natural capital and secure the future of humanity… where everybody has a role to play – from business to civil society, governments and local authorities, youth and indigenous people, private finance actors and development banks – to, altogether, reverse the trend of nature loss and preserve our natural capital (AXA and WWF, 2019).”

As for AXA, although the activities of the finance and insurance industry do not seem to be directly involved in biodiversity loss, its business relies heavily on the well-functioning of the economy which eventually depends on nature. The insurance sector has long been severely affected by catastrophes caused by natural disasters (Reuters, 2019). For example, such issues caused AXA 420 million euros more losses than usual in the second half of 2019. Thereout, AXA built a three-year partnership with the WWF on biodiversity and published the above joint recommendations (AXA, 2020). In addition to serving as a sustainable commitment, the partnership also reflects a growing recognition within the insurance industry of the financial materiality of ecological disruption and the need to develop precautionary strategies to mitigate the potential costs of future claims linked to nature-related disasters.

To attract more private sectors, this report warned that the biodiversity issue poses a real financial risk and sent a clear message: financial sectors cannot treat nature as an irrelevant matter anymore. Instead, they had to integrate nature-related risks into investment strategies, just like how they react to climate change. The joint report obviously targeted financial institutions as their main lobbying audience and framed biodiversity loss as a financial risk to make biodiversity more understandable to regulators and corporate managers. They are clear that economic actors will not invest any attention until the problems become unrelenting and costly (Ansari et al., 2013). This strategy is crucial in securing early attention and legitimacy from financial institutions and regulators who are primarily interested in and used to managing financial risks rather than ecological risks. As suggested by Arjaliès and Gibassier (2022), investors need to believe in the actual economic benefits of biodiversity before funding and acting. For example, the CEO of AXA Group explained the reasons for their participation:

“The potential loss of key ecological services endangers not only populations but also certain businesses that depend on them and can therefore become a concern for investors… by analysing our exposure to biodiversity-related risks and opportunities in our insurance and investment activities… (This) is a first attempt to map existing initiatives and call for a new public-private collaboration (AXA and WWF, 2019).”

Thus, although WWF and AXA seem to have unreconciled interests in fundamental, they share the same goal of shaping the sustainability reporting field which motivates them to eventually work together. This successfully gave the TNFD an entry ticket to the sustainability reporting field. WWF France and AXA subsequently presented this report during the G7 ministerial meeting in May 2019, where elites from finance, policy and environmental agendas were gathered. They acknowledged that the mainstreaming of the TCFD among financial institutions and corporations has inspired more initiatives to address other environmental issues and recommended that G7 Environment Ministers integrate nature into investment strategies.

In the wake of advertising this report, the motivation to establish a task force against biodiversity issues has been initially catalysed by Global Canopy, the United Nations Development Programme (UNDP), the United Nations Environment Programme Finance Initiative (UNEP FI), and WWF as four founding partners. This composition is crucial to the development of the TNFD. Voluntary frameworks are usually created with limited formal authority, which means their legitimacy highly depends on the knowledge of framework setters in given areas (Ahrne et al., 2007). As supernational bodies and well-recognised NGOs, their reputations and global influences were powerful weapons to attract some of the biggest players across diverse industries including finance, real estate, mining and agriculture to join the formation of the TNFD. Moreover, their join-in provided the TNFD with a high level of credibility (Cornish et al., 2023).

5.1.2 Defining the TNFD’s position: a flexible, market-led framework

When the influence of the matter becomes sufficiently problematic, or when voluntary framework evolves into standard expectation, government actions often follow (Österblom et al., 2022). This argument can also be observed in the TNFD’s case. At the end of 2019, France introduced a new decree under Article 29 of the Energy-Climate Law (French Government, 2021). By replacing previous Article 173, this newly published legal paper requires all French financial institutions to report their biodiversity-related risks and how they plan to control them. The new regulation that was about to be introduced could be seen as an overarching motivation for the French-based financial market to start thinking about biodiversity matters as they need to disclose their biodiversity risks from 2021. It established a formalised understanding of biodiversity as a compliance issue rather than an optional corporate responsibility. Ironically, the French government had already sought advice from some large multinational French financial institutions for the new reporting requirements before enacting them officially. By communicating with actors potentially affected by the law, the French government trying to make the regulatory process more effective (Luque-Vílchez et al., 2024). Nevertheless, most of the companies showed a negative response, not willing to follow the new decree. The intention behind their reaction is not hard to predict as companies are used to doing very little (Österblom et al., 2022) and rarely take full consideration of environmental issues (Elkington, 1994). Despite their initial resistance, many institutions later acknowledged the positive impact of such frameworks, illustrating that they needed regulatory intervention (AXA and WWF, 2019). This dynamic aligns with Maguire and Hardy's (2009) analysis, which shows powerful actors often engage in shaping emerging standards once they perceive change as inevitable. In this context, the shift of attitude of financial institutions can be read as seizing the opportunity and thereby influencing the further reporting landscape.

On this occasion, the absence of clear guidance on biodiversity disclosures left a strategic opening for the TNFD to position itself as the de facto standard setter. Although the TNFD recommendations had not been established at that time to ease immediate concerns about complying with the new regulation, according to the UNEP FI (2023a), it offers financial institutions a seat at the table during the framework development so they can affect the final recommendations towards the way they want.

Following that, at the TNFD's public announcement event, the French Secretary of State for Biodiversity endorsed the idea of establishing the task force:

"We are convinced that the work of the TNFD will accelerate the understanding of the (nature) issue. The TNFD is in line with the environmental strategy launched in France (Equator Initiative, 2020)."

However, in the same panel discussion, the only civil society representative from the Rainforest Action Network also raised some conundrums:

"There is a lot of potential for the TNFD; however, there is also a lot of risks. Research shows financial institutions are still fuelling harm with their investment and in fact disclose none of the harmful impacts. We really need to know where that money is going (Equator Initiative, 2020)."

A speaker from Global Canopy, one of the TNFD founding partners, immediately responded to this concern:

“Not all the money going out there into agriculture is nature negative. The problem we have is identifying what is bad and what is good (Equator Initiative, 2020).”

This response evaded the critical query raised by civil organisations and sought to construct the issue as a technical problem. By doing this, the TNFD shifted the issue from the harmful investment by private sectors towards definitional clarification. This is undoubtedly an artful way: It brought the NGOs to the negotiation table and acknowledged their concerns rhetorically but rarely translated them into agenda-setting. Under the appearance of participatory inclusiveness, financial actors still preserve their dominance on the stage.

With the intense advocacy, companies that perceive themselves as leaders recognise the benefits of active engagement and view such initiatives as vehicles for pre-competitive collaboration, bringing together firms that voluntarily commit to exceeding minimum legal requirements (Österblom et al., 2022). Leading influential financial institutions such as UBS, HSBC and BNP Paribas responded to the TNFD’s invitation by joining its membership and taking part in the pilot testing later on.

These efforts have yielded initial results, the UK Government immediately pledged up to £3 million, making it the largest donor to the TNFD. Initially, it considered launching a new UK programme that is parallel to the TNFD in order to anchor the UK’s Green Finance project. It soon realised more competing frameworks would dilute regulatory influence:

“Developing a parallel UK programme dilutes the convening power of the TNFD and is highly inefficient. It does not provide as good an opportunity as the TNFD to work closely with and learn from the TCFD given the TNFD already links up with the TCFD. It is unlikely Defra[4] would be able to create a global network of national governments and businesses in its own bilateral initiative (UK Government, 2021).”

This is because the existence of multiple equivalent rules would undermine their respective effects when standards are voluntary and all organisations are able to set new standards (Brunsson et al., 2012). Thus, it rooted for the establishment of the TNFD while giving up the dominant position and allowing private sector buy-in. The TNFD was then located as a finance-sector and corporate-led initiative on nature-based risk:

“This project will be the first example of a global collaborative approach to nature-based risk led by private financial institutions and the corporate sector, along with strong participation from central banks and regulators (Global Environment Facility, 2021).”

At the same time, the UK Government carefully avoided creating the impression that the TNFD is dominated by the UK financially or politically. Its effort to appear neutral

[4] Defra: Department for Environment Food and Rural Affairs, a UK government department.

can be seen as a way to protect the TNFD's image as a global and inclusive initiative. It thereby publicly emphasises the need for diverse funding sources:

"It is key that there is the right balance of contributions from all donors to ensure the TNFD is not perceived as a UK initiative…so it has independence from government and is genuinely private sector led (UK Government, 2021)."

Consistent with this objective, four founding partners announced the TNFD in July 2020 and an Informal Working Group (IWG) with over 60 members including financial institutions, corporations, governments and other influential actors. The composition of the IWG reinforces the narrative of broad and cross-sectoral involvement. It was temporarily set up to manage some initial work plans and decide on scopes for launching the formal task force in 2021. All IWG members work in designated groups and are assigned specific tasks (See Appendix 1). It strategically decided to use a phased approach which guides entities in progressively aligning with the framework over time. This staged method sets out three distinct stages of reporting requirements with increasing sophistication. Reporting users usually start with basic disclosures and gradually progress towards intermediary and eventually comprehensive alignment with the full framework. Thus, this approach clearly divided adopters into different compliance level by providing a structured pathway for engagement. The intention behind this clarity is to bring an image of order, allowing adopters to maintain a manageable sense of security when facing complex and unfamiliar problems. The TNFD also mentioned the plans to design reward mechanisms such as peer competing tools to motivate a higher level of compliance. However, on the other side, it might undermine comparability and defeat TNFD's intention to establish a consistent set of frameworks. A TNFD member organisation, BlackRock, expressed concerns that the overlaps and ambiguities between stages could lead to confusion rather than clarity (Zadek and Chambers, 2021). Notably, this stage approach was ultimately deleted from the final recommendations without any clarification or explanation.

5.1.3 Mythologising the legitimacy of its work

In January 2021, TNFD's three financial sector IWG Co-chairs [5]participated in the One Earth One Future Summit. One of the chairmen, from BNP Paribas, claimed that the TNFD enables investors, banks, insurers and companies to understand financial risks caused by nature degradation, and then integrate these risks into their investment, credit and insurance underwriting decisions (TNFD, 2021b). Meanwhile, the TNFD deliberately redefined nature loss as systemic financial risk in its Nature in Scope

[5] TNFD IWG Co-chairs:

- Antoine Sire, Director of Company Engagement and Member of the Group Executive Committee, BNP Paribas
- Dr Rhian-Mari Thomas OBE, Chief Executive, Green Finance Institute
- Mariuz Calvet Roquero, Director of Sustainability and Responsible Investment, Banorte

report (TNFD, 2021a), which directly affects asset values, investment portfolios and supply chains through ecosystem services. This report is the output of the informal working group, where the notion of materiality has been generally structured. The TNFD was imperious to achieve the crystallisation of meanings to obtain long-term legitimacy and broader implementation (O'Dwyer et al., 2024), as ambiguity might undermine institutional coherence and lead to misalignment (Feront and Bertels, 2019). Its materiality notion clearly consistent with the financial materiality approach adopted by the TCFD, requiring entities to disclose risks that both impact and being impacted by nature. Such a strategy attracted financial managers who were not interested in the biological risks. It linked the private sector's responsibilities with the TNFD objective, encouraging financial sectors to act:

“The financial sector risks inflicting significant damage on itself and companies across the world, if it fails to use its ‘great power’ to stop actions that harm the planet. As a core part of global economies, banks, investment firms and insurers, have a ‘great responsibility’ to ensure they support activities that are ‘nature positive’ and protect the air, land, water and animals in the Earth's unique and delicate ecosystems (EY et al., 2022).” The TNFD clearly constructed a logic that fits within the interests of financial institutions. It instrumentalises nature here and indirectly reflects its importance by measuring the risk of capital loss of economic participants.

Nevertheless, NGOs expressed their dissatisfaction greatly and critiqued such a financial-oriented approach:

“Under the TNFD's proposed framework businesses are fully entitled to continue to back the destruction and degradation of nature and human rights – so long as it doesn't affect their profits. But by not requiring companies to consider both long and short-term harms to communities and the environment in their risk analysis, the TNFD is setting the stage for severe environmental and human rights abuses and the worsening of the biodiversity and climate crises (Forests & Finance, 2022).”

According to Young (2006), standard setters need to make trade-offs and focus on the actors they value. These critiques reflected deeper concerns regarding the TNFD's focus on the private sector. As always, the TNFD had not directly responded to this complaint but accelerated its efforts to gain legitimacy through imitation, which can be considered the most remarkable feature of the TNFD. For example, the TNFD were acutely aware of the TCFD's widespread influence and recognised that aligning with the TCFD could significantly support the diffusion and legitimacy of the TNFD. As claimed by Rhian-Mari, another IWG co-chair from the Green Finance Institute, the TNFD is more like a biodiversity version of the TCFD and would adopt the same four-pillar framework as the TCFD, ensuring it as easy as possible for businesses and financial institutions to address climate and biodiversity risks simultaneously. Moreover, the name TNFD itself follows the TCFD name format that is intended to have brand associations and copies all four TCFD pillars. From the very beginning, they set the tone for mimicry of the successful work done by the TCFD. This attaches the TNFD to a rational, inevitable and even heroic appearance. Through these myth-making efforts, the TNFD created a perception that its emergence was both inevitable and desirable. Therefore, the TNFD situates its framework as the next phase in a broader programme of nature-related financial disclosures by linking itself to the TCFD's success:

"Key to the longevity and sustainability of the TNFD is that the Framework will be complementary to the TCFD. The IWG and Secretariat will seek alignment with TCFD processes wherever possible and eventually, the TNFD may seek integration with the TCFD in the future to create a comprehensive framework for environmental risk disclosure in the finance and corporate sectors (Global Environment Facility, 2021)."

It is apparent that the TNFD attempts to strengthen legitimacy, borrow credibility, and deify itself as a saviour for supporting organisations in tackling biodiversity issues by imitating the TCFD. To ensure further communication, a series of informal dialogues with the TCFD had been conducted, which affected the early thinking on the TNFD design. Nevertheless, the early idea was that the TNFD could be housed under the TCFD, as a "bolt-on". The World Bank Group also set out the rationale for such an idea:

"The initiative, which can be built on or be part of the TCFD, will provide a framework and guidance for regulating and supporting biodiversity reporting and risk assessment by real and financial sector firms. A TNFD framework can help avoid excessive additional requirements for real and financial sector firms and fragmentation of reporting standards (2020)."

However, the TCFD appeared reluctant to adopt this proposal (Cornish et al., 2023), possibly due to concerns that it might detract from its primary focus on climate issues (Cornish et al., 2023). Thus, the TNFD was created as an independent initiative and has deliberately avoided including specific climate disclosure recommendations in its later drafted report (TNFD, 2022).

Consequently, the TNFD was built as an international agenda more comparable to climate. Through framing itself as the nature equivalent of the TCFD: just as climate risks need a reporting framework, so do nature risks. The TNFD Co-Chair, David Craig, admitted that it combined the well-received TCFD because of familiarisation and lowering the friction, and they see the possibility to deal together the climate and nature issues (Tett, 2023). By framing itself as the nature equivalent of the TCFD, it creates normativity for its work: just as climate risks need a reporting framework, so do nature risks. It echoed the motivations of the sustainability officer at AXA during the TNFD launch event:

"My experience with the TCFD framework, which accelerated climate risk integration, gives me hope that the TNFD can do the same for biodiversity, providing necessary metrics and frameworks for investors to assess nature-related risks and opportunities (TNFD, 2021c)."

In general, this section summarised key events before the TNFD was officially established, starting from the WWF's call for a cooperative action to form a global initiative to the mobilisation of financial sectors by framing nature-related issues as a financial risk. It also shows that the earliest idea of imitating the TCFD has begun to take shape.

5.2 Design Phase

5.2.1 Iterative feedback loop and relation dynamics: inviting actors to defining its framework

The TNFD was formally launched with the endorsement of finance ministers from the Group of Seven (G7) under the leadership of Co-Chairs David Craig, and Elizabeth Maruma Mrema in June 2021. As a founder of the data and technology platform provider Refinitiv and a financial industry advisor, Craig's extensive experience and strong professional network helped bridge the gap between technological innovation and financial markets. Mrema is the Assistant Secretary-General of the United Nations and Deputy Executive Director of UNEP. Her leading role and personal connections within UNEP strengthened communication between the TNFD and UNEP and supported the TNFD to align with broader UN sustainable development agendas. Thus, the dual leadership of the TNFD allows both market perspectives and policy insights. However, on the other side, both the leadership and membership structures are fatal frustrations for the NGOs. While restoring nature and biodiversity are core objectives of the TNFD, the focus is deviated towards metrics, compliance and investor expectations:

"TNFD is just the latest in a succession of business-led initiatives that repeat a familiar pattern: it addresses biodiversity as an economic asset, adding a monetary value to it and quantifying it in market terms as "natural capital" stocks. It neglects the intrinsic value it has for the livelihoods of all people on this planet, negotiating its survival from a trading perspective (Forests & Finance, 2023)."

In this stage, the TNFD advanced a market-driven open innovation approach. The initial approach, as outlined in the Project Document in 2023, was planned to release a draft for consultation over 60 days before finalising the framework just like the predecessor TCFD. However, the TNFD Secretariat decided to adopt a more iterative and experimental open innovation method. Instead of a single consultation round, the TNFD released four successive beta versions (Beta v0.1 to Beta v0.4) over two years, each version incorporating public feedback from the previous round of consultation. The rationale behind this shift is that complexity of nature makes it difficult to directly borrow and transfer climate risk reporting (Cuckston, 2024). Such unfinished fashion extends and builds on the advocacy work in the preparatory stage, and the continuous feedback loop allowed the framework to evolve dynamically in response to real-time input from the perspectives of diverse actors. This is crucial to stabilise contested understandings and build sustainable coalitions (Rao et al., 2000). The iterative mechanism also reflected an intention to cultivate a sense of collective ownership and to build a sense of community and momentum. In the ecosystem field characterised by flux and controversy, it requires active maintenance and repair (Thomas and Ritala, 2021). By embedding ongoing consultation into the creation process, the TNFD blurred the boundaries between creating and maintaining institutional work (Lawrence and Suddaby, 2006), illustrating how defining work operates dynamically across different stages of institutionalisation.

Although the TNFD have included a wide range of actors in its framework-setting process, most of them played a consultative role rather than directly participating in decision-making like the TNFD members. All the 40 TNFD members are from financial institutions, corporates or market service providers. These strong market representatives are either veritably interested in bringing sustainability concerns into corporate reporting or they are stoutly to control an initiative that could probably threaten their extant positions (Flower, 2015). In the meantime, NGOs, civil organisations and indigenous representatives were relegated to the margins of the process, owning limited ability to influence decisions on core agendas. They clearly recognised that the TNFD is no different from previous initiatives under its guise of a multi-stakeholder method:

“On launch day we can expect TNFD to say that it has been ‘influenced’ by the views of civil society organizations and Indigenous Peoples. What it won’t do is respond to the substantive points that have been raised again and again on issues like complaints reporting, on lobbying, on the need for unilateral reporting on impacts (Rainforest Action Network, 2023).”

As for the iterative release of beta versions, the TNFD selectively incorporate controllable suggestions such as technical clarification and definitional refinements, while leaving more contentious issues unaddressed. Even though the TNFD has always presented itself as a multi-stakeholder initiative, its priorities remain firmly rooted in investor-oriented logic. NGOs’ requests seem being voiced through formal submissions, public forums and direct engagements during the beta consultations. However, while the TNFD Secretariat acknowledged the presence of such views, it did not structurally embed them in the final disclosure recommendations. The TNFD co-chair, Craig, also expressed that the TNFD cannot solve all issues that NGOs expected:

“TNFD can't solve every problem and every grievance. Obviously, we attract a lot of attention with our success, but there are a lot of grievances and issues out there that, frankly, TNFD will not solve, like land ownership rights and Indigenous peoples' rights. Those are recognised in the TNFD approach, but we can't solve all the grievances and issues. So, we've just got to be careful of how much people expect TNFD to do (See, 2025).”

Instead, the TNFD subtly adjusted its language, such as adopting the phrase “provide decision-useful information to capital providers and other stakeholders”, giving the appearance of greater inclusivity while keeping the core focus of market priority. This allows the TNFD to appear responsive to criticism without making any meaningful changes to its fundamental approach.

5.2.2 Balancing between ambiguity and clarification in its framework

The above approach seems to introduce greater uncertainty to the final outcome since the reiterative process is more flexible and adaptive. During this process, the original motivation of the TNFD for greater clarity seems watered down. As mentioned earlier, the TNFD was seeking to apply a sophisticated staged method which divides the organisation's compliance level into three different grades and devise incentive mechanisms such as peer group comparison tools to encourage higher-level adoption. However, this might greatly dampen the initial interest of institutions in joining the TNFD and exert tremendous pressure. In addition, companies also cannot rely too much on specific guidance considering the uniqueness of each organisation and the relevance of sustainability issues can change over time (O'Dwyer, 2011). Thus, the TNFD was then expected to bring a minimum of common requirements. In its Beta v0.2 version (2023), it put forward a common global set of core disclosure metrics for businesses and financial institutions across all sectors to ensure comparability. Such a certain degree of interpretive flexibility and ambiguity could defer conflicts between actors without requiring a unanimous consensus (Van Wijk et al, 2013; Ferraro et al., 2015). This is particularly crucial when reporting practices are still evolving.

The greater ambiguity is reflected not only in how the impact is structured but also in how implementation is required. The above core metrics are followed on a comply or explain basis, and not all 14 TNFD recommendations (Appendix 3) should be aligned to become a TNFD adopter. In fact, companies can still claim they are TNFD adopters by only following just one core recommendation. Furthermore, they reserve the discretion to determine which parts of their organisation the framework is applicable to, allowing selection or symbolic adoption. This high degree of flexibility offers clear benefits for firms. It reduces entry barriers and creates space for initial participation. By ensuring a larger extent of choices, actors would feel less pressured to take part. This is important because these early adopters will produce tangible products (i.e. the TNFD reports) that allow those who are still hesitating to see and replicate (Higgins et al., 2014). Such an idea was also confirmed in a TNFD-hosted webinar:

"The entry point into the TNFD and the requirements of an initial disclosure are deliberately set very low (Green Finance Institute, 2024)."

Meanwhile, an actor admitted the flexible approach encouraged initial attempts through its comment letter:

"The Taskforce's choice to use the phrase "adopting TNFD" rather than implementing, using, aligning to, complying with and various other options is superbly helpful to get organisations starting. Amongst our clients, at least, this flexibility is a clear incentive.

At the same time, it is not seen as a get-out clause. This is a major success of the TNFD. (EFTEC, 2024)"

While this approach allows greater flexibility, NGOs criticise that the TNFD had always selected metrics with a very low bar and did not publish meeting minutes, comments and survey responses:

"In its broader work, TNFD has often cherry-picked from approaches that set a very low bar, rather than reflect expectations set through more sound multi-stakeholder processes. Poor guidance is highly likely to set a lower bar than many existing national and international environmental and human rights standards. In doing so, TNFD will undermine the headway made, and lessons learned, over many years (Forests & Finance, 2022a)."

Moreover, they pointed out that the TNFD Secretariat verbally assured them that the framework would explicitly and clearly distinguish between "positive" and "negative" impacts on nature, which would allow for greater transparency in assessing corporate behaviours. However, in its final version, the TNFD discarded this commitment and instead adopted a "net" reporting approach, which means the "negative" impacts could be offset by good operations. The shift fundamentally altered how biodiversity impacts are measured, making these private sectors' true ecological influences invisible. The reasons behind this shift are not hard to understand. While whether to disclose all negative effects on nature is decided by a room full of corporations and financial institutions, the answer is unlikely to be positive. As satisfying one demand means sacrificing another (Pache and Santos, 2010), it suggests that the balance of the TNFD was tilted towards the businesses when conflicts cannot be easily reconciled.

The development of materiality is always a source of confusion (Jørgensen et al., 2021) and hard to construct precise definition (Edeley et al., 2014). Furthermore, O'Dwyer and Unerman (2020) indicate the inevitable subjectivity within the materiality notion of sustainability issues such as climate risk. As mentioned earlier, the TNFD announced the preference of using financial materiality that is consistent with the TCFD as a result discussed by the informal working group. This method is more about activities that will affect financial performance and corporate value. The decision was largely shaped by earlier mimicry work, where the TNFD was trying to further imitate the TCFD's financial materiality model in order to borrow legitimacy. It therewith published Beta v0.1 in 2022. Within this original version, it reiterated to align with the emerging global standard developed by the ISSB and methods adopted by specific jurisdictions (TNFD, 2022):

"The TNFD recommends that organisations follow an enterprise value approach aligned with the global baseline standards under development by the ISSB and aligned with the relevant jurisdiction in which reporting is performed."

This early alignment can be framed as a pragmatic step towards harmonisation and interoperability, while it can also be considered as a legitimacy seeking-strategy. It intended to position the TNFD within a rapidly converging transnational reporting infrastructure. By invoking the authority of the ISSB and connecting with pre-existing standards, the TNFD was able to insulate itself from public criticism, especially the concerns over its financial materiality approach. Although accelerated initial acceptance, the financial orientation struggled to fully capture the unique complexities of biodiversity-related risks. One year later, in its beta v0.2 recommendation, the TNFD reflected on the feedback it received for the first version and expanded its approach by including a set of guidance to assess impacts on nature which may go beyond the ISSB's requirements (TNFD, 2022b). While the early draft adhered closely to a financial materiality perspective, the feedback suggests an increasing willingness to engage with broader impact dimensions, even if such impacts are still within the frame in relation to enterprise value:

"The TNFD framework recognises that consideration of both nature-related dependencies and impacts is required for a comprehensive assessment of risks and opportunities, and that impacts on nature become relevant to enterprise value when assessed over a future time horizon (TNFD, 2022b)."

Thus, rather than a full embrace of double materiality, the TNFD appears to apply a more prudent approach. This framing soon became a site of contestation. A joint open letter from 28 NGOs and civil organisations criticised the TNFD's exclusive reliance on financial materiality, complaining that such an approach marginalised broader environmental impacts and risks overemphasising short-term profitability at the expense of long-term sustainability (Forests & Finance, 2022b). This idea has also been expressed by scholars who contend no matter how ingenious the financial reporting system is designed, the mere financial materiality cannot reflect all the implications of a company's activities (Bebbington et al., 2020; Laine, 2024). These critiques revealed an inherent tension in the TNFD's formation. While financial actors sought minimum effect on their operations, NGOs pushed for a broader materiality consideration. Even within the TNFD itself, some members such as BNP Paribas advocated for an expanded materiality perspective, contending that the enterprise value approach was inadequate to solve the most significant financial risks we experienced:

"A company's impacts to nature, for example, will not always create a foreseeable risk to that company, but may exacerbate the systemic risk of nature loss, which affects all companies. The complexity and severity of this systemic risk is entirely lost by placing enterprise value – as opposed to biosphere integrity - at the centre of concern (BNP Paribas, 2022)."

Instead of directly responding to this doubt, the TNFD co-chair, Craig, was taking a tough stance when referring to the criticisms of inclining to private actors:

"There is a misunderstanding about how we operate in this criticism. We have always been very clear that we are market-led. We are not changing that. The task force is chaired by myself, Razan (Al Mubarak) and its market members (See, 2025)."

Faced with these divergent voices and multiple pressures, the TNFD has to recalibrate its approach. This is because they realised that the recommendation cannot happen without immense cooperation between actors at the key notions (Hancher and Moran, 1989). In its final recommendation, the TNFD allows organisations to disclose through two materiality lenses: the financial significance criteria consistent with the ISSB and the broader materiality method that is used by GRI Standards (2023). This means organisations could choose the threshold of disclosing according to their needs. Such a mechanism defers conflict and enables gradual consolidation. It reveals how the TNFD actively maintained its framework while creating the final content. Thus, the dual method can be considered as a comprise or balancing move attempts to accommodate divergent interests.

5.2.3 Narrowing the criteria: educating and routinizing through industrial piloting testing

As the TNFD released its first beta version, the framework began to enter a more public-facing participatory phase, inviting market actors to engage directly with the draft and provide feedback. However, the response to some of the pilots revealed an early contradiction between the TNFD's objectives and the actual perspectives of potential adopters. According to over 500 pieces of feedback received in Beta v0.1, around a quarter of the feedback questioned piloting the framework is 'not that useful' (TNFD, 2022b), with some arguing that the framework's concepts were too abstract, while others cited insufficient internal capacity or limited access to necessary data and expertise (TNFD, 2022b). This scepticism was especially evident among organisations from Asia and Latin America, where language barriers and lower levels of existing nature-related reporting practice further constrained meaningful participation (UNEP FI, 2023b). According to the TNFD's news release on early adopters, only 6 per cent of organisations are based in Latin America and the Caribbean while the data for Africa and the Middle East was only 3 per cent (TNFD, 2024a). Although these regions are rich in natural and biodiversity resources, they were played down in the initial TNFD campaign, and even had no clue about what the TNFD is:

"There is a noticeable unfamiliarity with the Taskforce on Nature-related Financial Disclosures among Brazilian FIs, it is important to clarify that these guidelines have not been adopted in the country's regulatory frameworks (ERM NINT[6] and Global Canopy, 2023)."

"The Colombian financial sector is not yet familiar with the TNFD framework. The disclosure of information related to nature by the sector is not regulated (ERM NINT and Global Canopy, 2023)."

[6] EMR NINT: NINT is a consultancy in Latin America focused on sustainable finance and ESG advisory, acquired by the global sustainability consultancy ERM in 2023.

One prominent concern across emerging markets was the overlapping ESG frameworks. As highlighted in the UNDP's emerging market study, disclosure of nature-specific risks or impacts has not been mandated in many South American and Asia countries (UNDP, 2022). Without strong regulatory backing, companies remain hesitant to commit significant resources to voluntary disclosures. They were also getting fatigued and confused about how the TNFD disclosures interact with existing commitments to the TCFD, GRI and other frameworks:

"Some governments point to being overwhelmed by the rising number of ESG reporting and disclosure instruments recently, noting they are in constant need of improving their understanding and ability to deploy them. Making disclosure mandatory in a country requires time, capacity, and willingness. It is complicated for governments to bring these criteria each time a new type of reporting is introduced (UNDP, 2022)." However, on the other side of the earth, financial institutions also struggled with the data paucity and how the TNFD can be linked to other global frameworks and standards.

To mitigate these challenges, the TNFD Secretariat started to expand its educating work to accompany the piloting process. The TNFD recommendation materials have been translated into six widely used languages, A series of online webinars, case studies and guidance materials were introduced, aimed at bridging the knowledge gap and interpreting the TNFD framework in understandable ways. The co-chair, Craig, focusing the next stage on educating companies:

"The other thing we learned very quickly is great frameworks are nothing without adoption…There was no training education or capabilities. I feel like we invest so much energy debating taxonomies but we don't invest enough in training and education. So, we've embarked on the whole workstream about capability building (Tett, 2023)."

This section discussed four beta stages of the TNFD's framework before officially publishing its final recommendations. While NGOs pushed for stronger transparency of disclosure and more rigorous standards, their voices gave way to corporate interests. This was reflected in the shift to "net" impact reporting and a minimal compliance requirement. Meanwhile, the development of the materiality notion is also an emphasis in this stage, which initially centred on financial materiality but gradually expanded to incorporate factors of double materiality under pressures from public feedback. During the evolving process, companies around the world faced frustration to a different extent.

5.3 Adoption phase

5.3.1 Embedding itself into existing reporting practice

Despite the flexibility in coordinating collaboration among different participations and bridging their divergent interests, its double-edged nature may also bring interminable disagreements (Fischhendler, 2008). Even though some actors may collectively agree on a settlement, it does not mean they will regard this as a final decision (Litrico and David, 2017). Therefore, the imperative to secure its status among numerous global framework setters became particularly apparent when the TNFD published its final recommendations in 2023 and moved into the adoption stage. At this stage, the focus of the task force shifted from developing guidance to aligning with the prevailing standard setters. This reflects that the TNFD attempts to send a signal to regulators and the public that it can be used as a complementary framework rather than competition. For example, the TNFD has jointly published a mapping of the correspondence with the European Financial Reporting Advisory Group (EFRAG), responding to the requirements of market participants under the Corporate Sustainability Reporting Directive (CSRD). A high level of commonality was achieved in the definition, materiality approach and the adoption of the LEAP method[7] (EFRAG, 2024). Subsequently, these two organisations signed a memorandum of understanding (MoU) that reduced the reporting burden for reporting entities and further consolidated the mimicry strategy. As claimed by the EFRAG Chair Cambourg, this indicated a significant milestone in promoting transparency in nature-specific corporate reporting:

“As Chair of the EFRAG SRB, I am pleased to see the culmination of our collaborative efforts with TNFD in maximising consistency and providing clarity for market participants. This mapping will practically support preparers and other stakeholders to understand how to leverage on TNFD when reporting on ESRS and vice versa. Moving forward, EFRAG remains dedicated to supporting the implementation of nature-related disclosures and collaborating with TNFD to develop further guidance and tools for market participants (EFRAG, 2024).”

Meanwhile, the TNFD has deepened its cooperation with the GRI and co-wrote a report to reinforce the TNFD’s claims of consistency and complementarity. It enables companies to comply with the existing reporting requirements while gradually incorporating biodiversity considerations. The TNFD was trying to embed itself into the existing reporting practice and reduce the friction of adoption. Under the purpose of harmonisation, what we may be witnessing is voluntary frameworks subtly shaping the interpretation and implementation of mandatory standards. However, this may blur its distinctiveness, which lies in its multi-stakeholder focus and in its development of new tools such as the LEAP method. As the TNFD increasingly integrates with other frameworks and regulatory regimes, the boundaries between its unique contribution and pre-existing standards become less clear. This makes it more difficult to justify the TNFD’s separate existence and articulate what it offers beyond existing disclosure norms.

[7] The LEAP Method: It is an integrated approach for the assessment of nature-related issues that is designed for use by organisations of all sizes and across all sectors and geographies. The LEAP method refers to the short of locating, evaluating, assessing and preparing.

5.3.2 Constructing the regulatory-oriented identity

On the other side, the TNFD has set its sights on connecting with global biodiversity goals. Given the persistent concerns around its credibility, authority and enforceability that are rooted in its voluntary nature, the above strategy can be understood as a deliberate effort to gain output legitimacy. As explained by Suchman (1995), the legitimacy threshold would be stringent if an organisation aspires to active support from a wide range of actors. In this light, the strategic alignment reflects a deep sense of insecurity. The absence of government regulations, contested participation requirements and the flexibility in materiality method and critique of the inclusiveness of minority representatives increased its vulnerability, forcing the TNFD to align with powerful global reporting regulations to compensate for such limitation. It is indispensable for the survival of voluntary initiatives such as the TNFD.

This intention can be further perceived at COP 16 in Colombia in October 2024, where the TNFD released a discussion paper on a nature transition plan, directly linking its framework to the Kunming-Montreal Global Biodiversity Framework (GBF). By positioning itself as a key enabler of the GBF's ambitious goal, the TNFD seeks to frame its identity as the vehicle to align GBF's ambitions within corporate strategy. This approach attempts to enhance its legitimacy by placing its framework with normative authority and softly transferring itself as an implementation tool of global policy. However, while linking to the GBF may increase perceived relevance, it does not resolve foundational critiques about the TNFD's transparency, inclusiveness and power asymmetry.

At the adoption stage, the TNFD's focus has shifted to align with other reporting initiatives and global standards. Although it could be considered a significant strategy to increase its status and perceived legitimacy, more evidence should be collected about its implementation as it is still evolving.

6. Discussion

This study reveals how core tensions unfolded during the emergence of the TNFD, and how participating actors simultaneously engaged in creating and maintaining institutions. By narrating the dynamical forms of institutional work across the preparatory, design and adoption stages, this study uncovers how actors constantly adapted their strategies to secure their legitimacy, mobilise supports and negotiate in a fragment nature-related risk reporting field. Drawing on Lawrence and Suddaby's (2006) typology theory, it shows that the institutional work in this TNFD's case did not follow a linear process but instead evolved through recursive interactions between different forms across the creation and maintenance state during the four beta stages. For example, creating work such as advocacy, defining and mimicry were not confined to the early stage of the TNFD but continue to resurface in later phases, especially in response to external questioning and requirements. At the same time, maintenance work such as embedding and routinising keep developing and bringing innovation to the TNFD framework. This iterative entanglement provides empirical evidence of the interweave of institutional work in emergent framework setting process. To further figure out the complex dynamic, this paper introduces a functional categorisation of institutional work observed in the TNFD process: (1) legitimacy building, which sought to construct external legitimacy and position itself within the nature reporting landscape; (2) incremental permeation, which aimed to embed its framework within the routines of actors' daily operation; (3) coalition creation, which focuses on assembling an alliance of governments, NGOs, corporates and financial institutions, although they have divergent interests and expectations.

This section starts by tracing the dominant institutional work that is used in the three chronological phases, from preparatory mobilisation to framework design and the ultimate adoption stage. Then, it illustrates how these work types were recombined into the three functional categories that support the TNFD's institutionalisation. It highlights how different works interact, reinforce and contradict each other. Lastly, it also reflects the shift from providing great level of clarity to allowing more ambiguity and flexibility in its framework.

6.1 Relations and dynamics

This research extends Lawrence and Suddaby's (2006) institutional work taxonomy by illustrating how different forms of institutional work interact dynamically through the TNFD's formation. Unlike some early studies that focus on isolated work forms (Zietsma and McKnight, 2009; Trank and Washington, 2009), it echoes recent papers (Canning and O'Dwyer, 2016; Micelotta et al., 2013) and treats the framework setting process as recursive, overlapping and characterised by iterative adjustments. The TNFD's institutionalisation work indicates the difficulties in constructing a reporting framework amidst heterogeneous actors. This process is not merely technical but involves various political and social skills at different stages (Garud et al., 2002). Building on the preceding findings, this research generates three distinct but interrelated functional categories from the TNFD's emergence:

Legitimacy building comprises works designed to construct an impression that insulates institutions from external pressures and questions (Meyer and Rowan, 1977). The TNFD's initial work centred on creating demand and exerting influence by stressing the urgency and importance of nature issues. It problematised nature loss in financial terms, which can be read as a deliberate act that defines ecological degradation as a threat to enterprise value. Thus, instead of presenting biodiversity decline as a conventionally social or ecological crisis, it situated abstract ecological considerations clearly within the market logic, enhancing the awareness of powerful regulators and investors. This decision was a targeted advocacy activity aimed at financial institutions and corporations who might otherwise view biodiversity as peripheral or unquantifiable. This idea was further amplified by the endorsement of governments such as the UK authority which helps enhance the legitimacy of the TNFD's voluntary framework. To reinforce the above strategy, the TNFD engaged with mimicry work that stressed nature issues as the counterpart of climate risks. By adopting the same four pillars and financial-oriented materiality method as the TCFD, the TNFD further creates a sense of mission for financial institutions and resonates emotionally with it. In this early phase, the TNFD demonstrated a strong tendency towards clarity. This is an inevitable outcome for achieving high consistency with the TCFD. However, while winning the favour of financial institutions, other participants under its multi-stakeholder focus are bound to be dissatisfied that tensions would later surface as the framework develops.

The combination of educating, embedding and routinising work constitutes what is conceptualised as the process of *incremental permeation*. As the TNFD entered its design phase after its official establishment, it confronted intensifying tensions that exposed the limits of its early clarity. The critique from NGOs and civil organisations on the low bar metrics and financial materiality has increased. These actors called for greater emphasis on biodiversity and wider stakeholder interests. Despite this, the TNFD responded by loosening its commitments in place of directly addressing these concerns. For example, the TNFD transited from the previously promised distinction between "positive" and "negative" nature impacts to a more ambiguous "net" impact. This shift added interpretive flexibility that allows negative effects to be offset through compensatory actions. This method of deliberately introducing more ambiguity conceals companies' harmful activities, making the contradictions less prominent. Meanwhile, the application of ambiguity was further developed through the rhetorical inclusion of its materiality approach. While early drafts focus on relatively narrowed financial materiality, the final version confirmed a pluralistic setting that allows entities to choose the materiality logic that fits their needs. The double lenses offer discretion for organisations to meet jurisdiction requirements but they also leave options for actors who prefer investor-oriented minimum disclosure. This ambiguity helped heterogeneous actors integrate their own priorities into the framework. In parallel, the TNFD conducted a series of educating, embedding and routinising work, primarily through pilot testing with institutions across regions and sectors. These events aimed at capacity building, attempting to reduce knowledge gaps, especially for areas that are unadapted to biodiversity metrics. The TNFD was able to foster familiarity and expand its usage without intensifying tensions.

The *coalition creation* involves mythologising and identity construction which shapes institutional distinctiveness while reconciling interests with existing competitors. In its later phase, the TNFD's working emphasis transferred from constructing its framework to embedding itself with extant global reporting initiatives and aligning with high-level sustainability goals. The interoperable metrics signed with GRI and EFRAG reduce uncertainty for initial participants and produce formal association. Beyond that, the TNFD also sought to raise its publicity by linking itself to global policy agendas, such as the Kunming-Montreal GBF. Through the release of its “nature transition plan” at COP 16, the TNFD mythologise its role as a vital translator bridging global biodiversity goals and corporate reporting practices. In this way, it had constructed a mission-filled identity that dilutes the early criticism of its investor orientation. This strategy of seeking flexible alignment supports the TNFD to be adopted widely without mandates or regulations. The below process model explains such categorisation and how they interact with each other (Table 3).

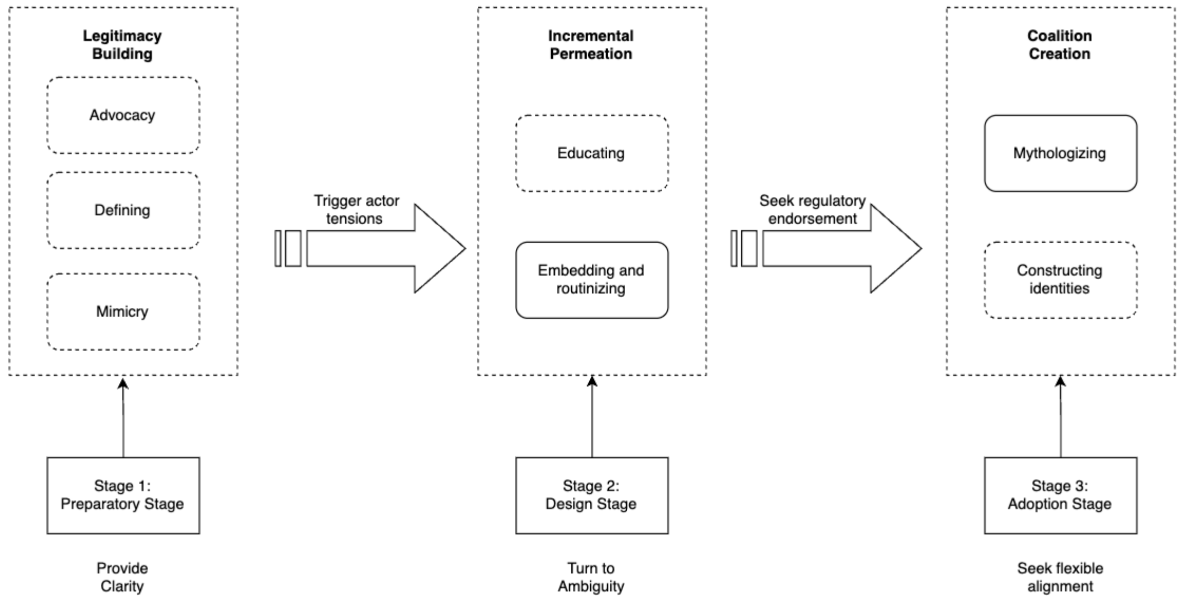


Table 3: Theoretical framework (created by author)

---------: Creating institutional work

———: Maintaining institutional work

1. This model illustrates the mechanisms of institutional work in the TNFD process, highlighting the dynamic interaction between institutional creation and maintenance.

2. The categories (legitimacy building, incremental permeation, coalition creation) represent different strategies employed by actors to establish and sustain institutions.

7. Conclusion

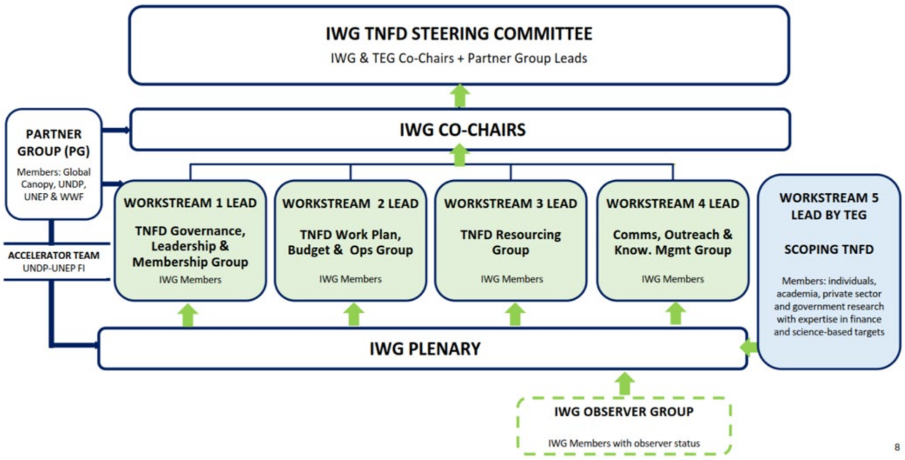
This paper investigates the emergence of the TNFD, focusing on the tensions that occurred and how the TNFD have managed these contradictions through diverse strategies. This paper employs the institutional work taxonomy (Lawrence and Suddaby, 2006) to analyse how different work forms have promoted this process. This extends the previous standpoints which suggest the institutional work should be more flux and the boundary between creating and maintaining institutions is not necessarily clear (Canning and O’Dwyer, 2016), examining the dynamics and interactions within the biodiversity framework setting. A process model was created, describing how they interrelated over time and work towards three distinct shared functions.

This research contributes to the literature in the following dimensions. Firstly, it enriches a limited understanding of the biodiversity framework-setting process. While previous focus has been largely put on the climate aspect, there is relatively little research on nature-related financial disclosures. Although some papers have examined the emergence of sustainability reporting initiatives, including GRI and the TCFD, the diversity of participants in shaping the TNFD's recommendation makes this process chaotic and disordered. Investigating the successful story behind the TNFD's formation shows that this process is not smooth and sequential, but more like a recursive process which needs defining and redefining its identities through negotiation. Secondly, it demonstrates how different institutional work forms (Lawrence and Suddaby, 2006) across creating and maintaining institutions do not take place in isolation but interact to reproduce broader function categories and drive institutions emergence. Specifically, advocacy, defining and mimicry converge to fulfil legitimacy construction goals, while educating, embedding and routinising combined to incrementally permeate the TNFD framework into corporate daily management. And then mythologising and constructing identities work jointly to secure its status. This is not simply clustering and naming institutional work. It attempts to explain how they evolve, combine and promote institutionalisation at different stages. It means each phase of work reinterprets and reinforces endeavour from the last stage. This dynamic interplay of work forms provides a richer and processual understanding of how legitimacy is built in contested spaces. Lastly, this research also responded to the lack of field-to-field interactions (O'Dwyer, 2023; O'Dwyer et al., 2024; Zietsma et al., 2017) by unfolding how diverse actors such as NGOs and private financial institutions collectively conduct institutional work to establish a shared framework at the field level.

While this paper attempts to capture actor interactions and dynamics of the TNFD formation process, it heavily relies on public documents such as reports, news releases, pilot testing documents and conference transcripts. Although this research made significant efforts to cover viewpoints from key actors involved in the TNFD process, it is crucial to acknowledge that most of the available data is official information produced by the TNFD itself and its organisational partners. Meanwhile, the TNFD has not made most comment letters and meeting agendas public, limiting researchers' ability to fully unveil power struggles, and informal contradictions compromise that shaped its framework development. Although this research is triangulated across multiple sources to mitigate the potential bias, such as complaints from NGOs and social organisations, the absence of direct access to internal documents and members means that some of the contentions might be concealed or glorified. Thus, future research could address this risk by conducting closer contact with the TNFD and discovering micro-level interactions, such as accessing internal records and interviewing key actors. In addition, this research unveils the interactions between different forms of institutional work based on the formation of the TNFD. Although the TNFD is a great example situated at the intersection of sustainability and financial disclosures, the single case study would inevitably raise questions about reliance on context and transferability (Yin, 2018). Further research could employ or refine the process model into the analysing the development of other reporting initiatives and assess whether it is applicable under different phenomena. While at the time of writing, the TNFD is still in the middle stage of its ambitious expansion. Thus, future studies could pay close attention to how the TNFD align with accounting standard setters such as the ISSB and back-stage interactions between them.

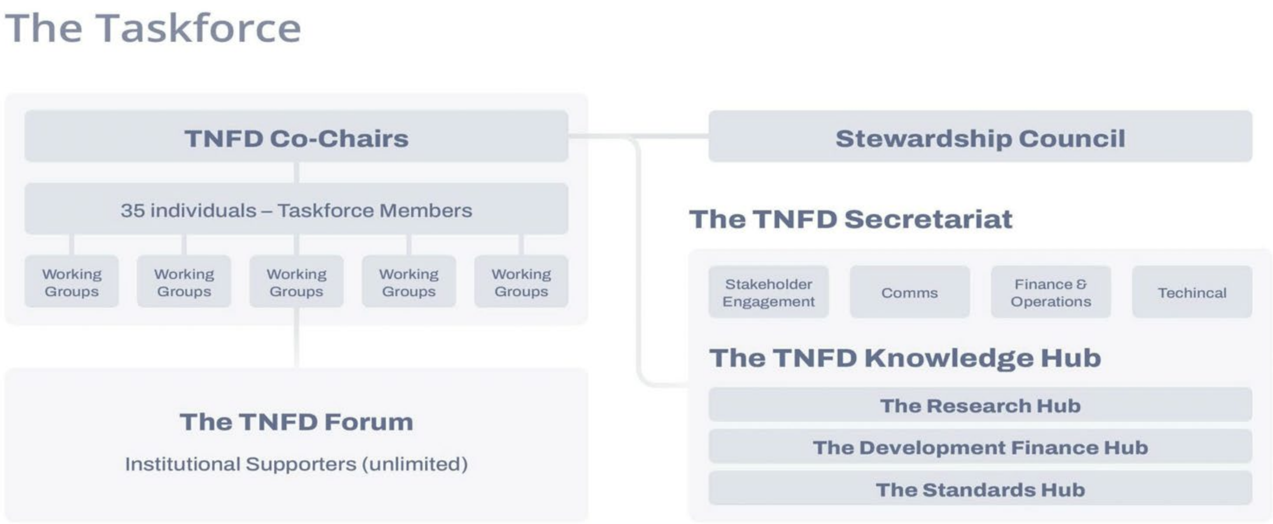
Appendix:

1. IWG Composition



(WWF GEF Project document, 2021)

2. TNFD components and Secretariat membership



(TNFD, 2021)

The Secretariat Leadership Team:

Tony Goldner: Chief Executive Officer

Emily McKenzie: Technical Director

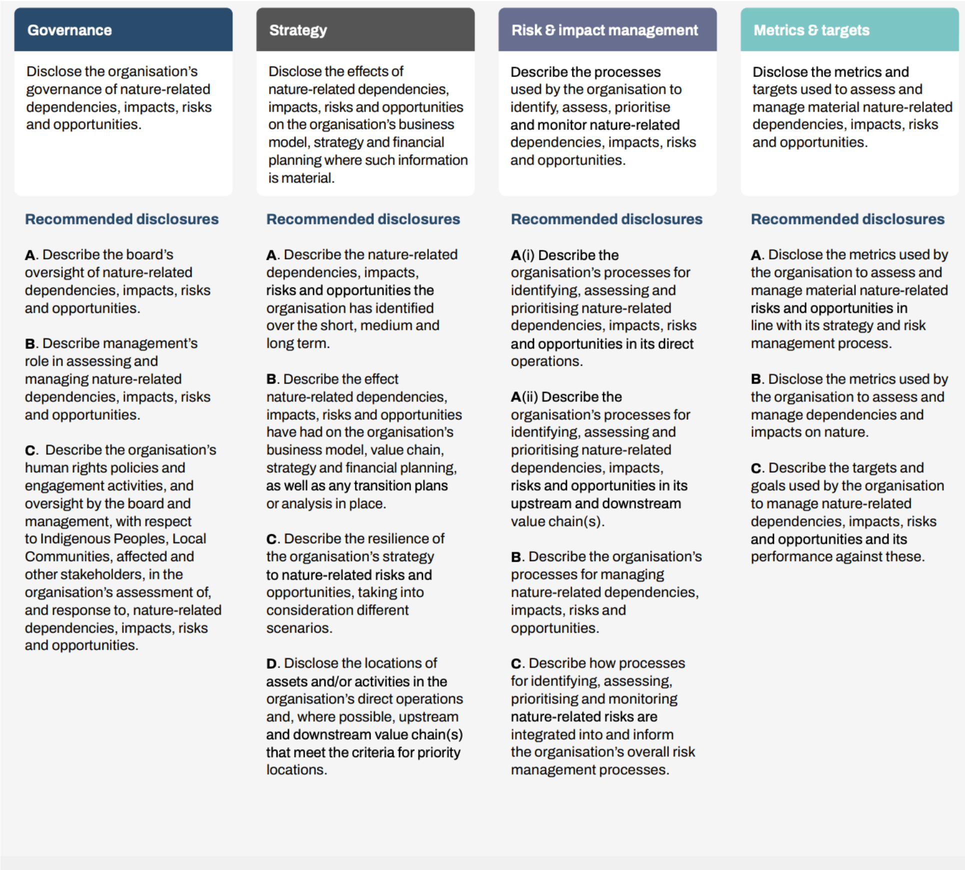
Barbara Sanderson: Chief Operating Officer

Candice Dott: Director of Market Engagement

Cathrine Armour: Director of Data Initiatives

Rita Lockheart: Communications Lead

3. TNFD recommendations:



(TNFD, 2023)

References:

1. Ahrne, G., Brunsson, N. and Hallström, K.T. (2007). Organizing Organizations. *Organization*, 14(5), pp.619–624.
2. Allini, A., Aria, M., Macchioni, R. and Zagaria, C. (2018). Motivations behind users' participation in the standard-setting process: Focus on financial analysts. *Journal of Accounting and Public Policy*, 37(3), pp.207–225.
3. Alon, A. and Dwyer, P.D. (2016). SEC's acceptance of IFRS-based financial reporting: An examination based in institutional theory. *Accounting, Organizations and Society*, 48, pp.1–16.

4. Alvesson, M. and Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review*, 32(4), pp.1265–1281.
5. Ansari, S., Wijen, F. and Gray, B. (2013). Constructing a Climate Change Logic: An Institutional Perspective on the 'Tragedy of the Commons'. *Organization Science*, 24(4), pp.1014–1040.
6. Arjaliès, D. and Gibassier, D. (2022). Can Financialization Save Nature? The Case of Endangered Species†. *Contemporary Accounting Research*.
7. AXA (2020). *AXA launches a partnership with WWF France*. [online] AXA. Available at: https://www.axa.com/en/news/AXA-launches-a-partnership-with-WWF-France [Accessed 9 Jun. 2025].
8. AXA and WWF (2019). *Into the Wild: integrating nature into investment strategies*.
9. Battilana, J., Leca, B. and Boxenbaum, E. (2009). How Actors Change Institutions: Towards a Theory of Institutional Entrepreneurship. *The Academy of Management Annals*, 3(1), pp.65–107.
10. Baudot, L. (2014). GAAP convergence or convergence Gap: unfolding ten years of accounting change. *Accounting, Auditing & Accountability Journal*, 27(6), pp.956–994.
11. Bellucci, M., Simoni, L., Acuti, D. and Manetti, G. (2019). Stakeholder engagement and dialogic accounting. *Accounting, Auditing & Accountability Journal*, 32(5), pp.1467–1499.
12. Blackstone, A. (2019). *Social Research: Qualitative and Quantitative Methods*. 2nd ed. FlatWorld.
13. BNP Paris (2022). *BNP PARIBAS RESPONSE TO ISSB CONSULTATION – Exposure Draft S1 General Requirements for Disclosure of Sustainability-related Financial Information*. [online] BNP Paris. Available at: https://cdn-group.bnpparibas.com/uploads/file/exposure_draft_ifrs_s1_general_requirements_for_disclosure_of_sustainability_related_financial_information_bnpp_response_issb_consultation_july_2022.pdf [Accessed 12 Mar. 2025].
14. Bowen, G.A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), pp.27–40.
15. Brint, S. (2020). *In an Age of Experts: The Changing Roles of Professionals in Politics and Public Life*. Princeton: Princeton University Press.

16. Brundtland Commission (1987). *Report of the world commission on environment and development: Our common future*. [online] UN. Available at: https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf [Accessed 2 Jan. 2025].
17. Brunsson, N., Rasche, A. and Seidl, D. (2012). The Dynamics of Standardization: Three Perspectives on Standards in Organization Studies. *Organization Studies*, 33(5-6), pp.613–632.
18. Burchell, S., Clubb, C., Hopwood, A., Hughes, J. and Nahapiet, J. (1980). The roles of accounting in organizations and society. *Accounting, Organizations and Society*, 5(1), pp.5–27.
19. Canning, M. and O'Dwyer, B. (2016). Institutional work and regulatory change in the accounting profession. *Accounting, Organizations and Society*, 54, pp.1–21.
20. Chiapello, E. and Medjad, K. (2009). An unprecedented privatisation of mandatory standard-setting: The case of European accounting policy. *Critical Perspectives on Accounting*, 20(4), pp.448–468.
21. Christensen, H., Hales, J., O'Dwyer, B. and Peecher, M.E. (2024). Accounting for sustainability and climate change: Special section overview. *Accounting, Organizations and Society*, 113, p.101-568.
22. Chua, W.F. and Taylor, S.L. (2008). The rise and rise of IFRS: An examination of IFRS diffusion. *Journal of Accounting and Public Policy*, 27(6), pp.462–473.
23. Clark, B.R. (1960). The 'Cooling-Out' Function in Higher Education. *American Journal of Sociology*, 65(6), pp.569–576.
24. Clune, C. and O'Dwyer, B. (2020). Framing Engagement that Resonates: Organizing Advocacy for Corporate Social and Environmental Accountability. *European Accounting Review*, 29(5), pp.851–875.
25. Cooper, D.J., Hinings, B., Greenwood, R., Brown, J.L., Cooper, D.J., Hinings, B., Greenwood, R. and Brown, J.L. (1996). Sedimentation and Transformation in Organizational Change: The Case of Canadian Law Firms. *Organization Studies*, 17(4), pp.623–647.
26. Cooper, S.M. and Owen, D.L. (2007). Corporate social reporting and stakeholder accountability: The missing link. *Accounting, Organizations and Society*, 32(7-8), pp.649–667.

27. Cornish, A., Champagne, J.-M. and Hilton, S. (2023). *MID-TERM REVIEW OF THE ESTABLISHING THE TASKFORCE ON NATURE-RELATED FINANCIAL DISCLOSURES (TNFD) PROJECT*.
28. Cuckston, T. (2024). THE TNFD: Reimagining corporate contributions to a ‘nature-positive’ future. In: A. Venturelli and C. Mio, eds., *The Routledge Handbook of Accounting for the Sustainable Development Goals*. London: Routledge.
29. Currie, G., Lockett, A., Finn, R., Martin, G. and Waring, J. (2012). Institutional Work to Maintain Professional Power: Recreating the Model of Medical Professionalism. *Organization Studies*, 33(7), pp.937–962.
30. Denzin, N.K. and Lincoln, Y.S. eds., (1994). *Handbook of qualitative research*. 1st ed. Thousand Oaks: Sage.
31. Diane-Laure Arjaliès, Laurel-Fois, D. and Mottis, N. (2022). Prison break from financialization: the case of the PRI reporting and assessment framework. *Accounting, auditing & accountability/Accounting auditing & accountability*, 36(2), pp.561–590.
32. DiMaggio, P.J. and Powell, W.W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, 48(2), pp.147–160.
33. Durocher, S., Fortin, A. and Côté, L. (2007). Users’ participation in the accounting standard-setting process: A theory-building study. *Accounting, Organizations and Society*, 32(1-2), pp.29–59.
34. Dyer, W.G. and Wilkins, A.L. (1991). Better Stories, Not Better Constructs, to Generate Better Theory: A Rejoinder to Eisenhardt. *The Academy of Management Review*, 16(3), p.613.
35. EFRAG (2024). *TNFD and EFRAG publish correspondence mapping*. [online] EFRAG. Available at: https://www.efrag.org/en/news-and-calendar/news/tnfd-and-efrag-publish-correspondence-mapping [Accessed 12 Mar. 2025].
36. EFTEC (2024). *Comment letter to the TNFD Secretariat*. [online] Economics for the Environment. Available at: https://tnfd.global/wp-content/uploads/2024/05/20240328-eftec-Comment-letter-Additional-Guidance-for-Financial-Institutions.pdf [Accessed 12 Mar. 2025].

37. Elkington, J. (1994). Towards the Sustainable Corporation: Win-Win-Win Business Strategies for Sustainable Development. *California Management Review*, 36(2), pp.90–100.
38. Empson, L., Cleaver, I. and Allen, J. (2013). Managing Partners and Management Professionals: Institutional Work Dyads in Professional Partnerships. *Journal of Management Studies*, 50(5), pp.808–844.
39. Equator Initiative (2020). *Transforming Finance for Nature: A task force to support transparency and accountability for finance and nature*. [online] YouTube. Available at: https://www.youtube.com/watch?v=NDreE2FvNso [Accessed 1 Apr. 2025].
40. Etzion, D. and Ferraro, F. (2010). The Role of Analogy in the Institutionalization of Sustainability Reporting. *Organization Science*, 21(5), pp.1092–1107.
41. EY, Microsoft and Earth Knowledge (2021). *Financial sector has 'great power' to stop biodiversity loss and reap the rewards of being 'nature positive', report reveals*.
42. Farooq, M.B. and de Villiers, C. (2020). How sustainability assurance engagement scopes are determined, and its impact on capture and credibility enhancement. *Accounting, Auditing & Accountability Journal*, 33(2), pp.417–445.
43. Fereday, J. and Muir-Cochrane, E. (2006). Demonstrating Rigor Using Thematic Analysis: a Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*, 5(1), pp.80–92.
44. Fernández Chulián, M., Garcia-Torea, N., Larrinaga, C. and Bebbington, J. (2024). Boundary objects: sustainability reporting and the production of organizational stability. *Accounting, Auditing & Accountability Journal*, 38(2).
45. Feront, C. and Bertels, S. (2019). The Impact of Frame Ambiguity on Field-Level Change. *Organization Studies*, 42(7), pp.1135–1165.
46. Ferraro, F., Etzion, D. and Gehman, J. (2015). Tackling Grand Challenges Pragmatically: Robust Action Revisited. *Organization Studies*, 36(3), pp.363–390.

47. Fischhendler, I. (2008). When Ambiguity in Treaty Design Becomes Destructive: A Study of Transboundary Water. *Global Environmental Politics*, 8(1), pp.111–136.

48. Flick, U. (2004). Triangulation in Qualitative Research. In: U. Flick, E. von Kardoff and I. Steinke, eds., *A Companion to Qualitative Research*. London: SAGE Publications.
49. Flower, J. (2015). The International Integrated Reporting Council: A story of failure. *Critical Perspectives on Accounting*, 27, pp.1–17.
50. Forests & Finance (2022a). *Civil Society Groups Say UN-backed TNFD Proposal Risks Opening a New Frontier for Corporate Greenwashing on Nature*. [online] Forests & Finance. Available at: https://forestsandfinance.org/news/civil-society-groups-say-un-backed-tnfd-proposal-risks-opening-a-new-frontier-for-corporate-greenwashing-on-nature/ [Accessed 2 Apr. 2025].
51. Forests & Finance (2022b). *Joint NGO open letter to the TNFD, with feedback on their draft framework*. [online] Forests & Finance. Available at: https://forestsandfinance.org/wp-content/uploads/2022/05/Letter-on-TNFD-feedback-1-1.pdf [Accessed 12 Mar. 2025].
52. Forests & Finance (2023). *TNFD final framework launches to ongoing fears of greenwashing - Forests & Finance*. [online] Forests & Finance. Available at: https://forestsandfinance.org/news/tnfd-final-framework-launches-to-ongoing-fears-of-greenwashing/ [Accessed 8 Apr. 2025].
53. French Government (2021). *Publication of the implementing decree of Article 29 of the Energy-Climate Law on non-financial reporting by market players*. [online] Direction générale du Trésor. Available at: https://www.tresor.economie.gouv.fr/Articles/2021/06/08/publication-of-the-implementing-decree-of-article-29-of-the-energy-climate-law-on-non-financial-reporting-by-market-players [Accessed 12 Mar. 2025].
54. Garud, R., Jain, S. and Kumaraswamy, A. (2002). INSTITUTIONAL ENTREPRENEURSHIP IN THE SPONSORSHIP OF COMMON TECHNOLOGICAL STANDARDS: THE CASE OF SUN MICROSYSTEMS AND JAVA. *Academy of Management Journal*, 45(1), pp.196–214.

55. Gioia, D.A., Corley, K.G. and Hamilton, A.L. (2012). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, 16(1), pp.15–31.
56. Giordano-Spring, S., Larrinaga, C. and Géraldine Rivière-Giordano (2024). Field-configuring events and the failure to standardise accounting for carbon emissions. *Accounting Auditing & Accountability Journal*, 37(9), pp.216–247.
57. Global Canopy (2020). *Missing metrics: climate vs nature – Global Canopy*. [online] Global Canopy. Available at: https://globalcanopy.org/insights/insight/data-on-nature-related-risks-is-more-complicated-than-for-climate/ [Accessed 5 Jun. 2025].
58. Global Environment Facility (2021). *GEF-7 REQUEST FOR CEO ENDORSEMENT / APPROVAL CHILD PROJECT – MSP ONE-STEP*. [online] *Global Environment Facility*. Available at: https://files.worldwildlife.org/wwfcmsprod/files/Publication/file/77z1d42c3d_G0034_WWF_GEF_CEO_Endorsement_Request_Establishing_the_Taskforce_for_Nature_Related_Financial_Disclosures_final_050421.pdf [Accessed 1 Apr. 2025].
59. Gray, B. and Purdy, J. (2018). *Collaborating for Our Future: Multistakeholder Partnerships for Solving Complex Problems*. 1st ed. *Oxford Scholarship Online*. Oxford: Oxford University Press.
60. Green Finance Institute (2024). *UK action on TNFD*. [online] YouTube. Available at: https://www.youtube.com/watch?v=3Eb7riYc3ys [Accessed 12 Mar. 2025].
61. Greenwood, R. and Hinings, C.R. (1996). Understanding Radical Organizational Change: Bringing together the Old and the New Institutionalism. *The Academy of Management Review*, 21(4), p.1022.
62. Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E.R. and Lounsbury, M. (2011). Institutional Complexity and Organizational Responses. *Academy of Management Annals*, 5(1), pp.317–371.
63. Greenwood, R. and Suddaby, R. (2006). Institutional Entrepreneurship In Mature Fields: The Big Five Accounting Firms. *Academy of Management Journal*, 49(1), pp.27–48.
64. Gross, T. and Zilber, T.B. (2020). Power Dynamics in Field-Level Events: A Narrative Approach. *Organization Studies*, 41(10), pp.1369–1390.

65. Hancher, L. and Moran, M. eds., (1989). *Organizing Regulatory Space*. 1st ed. Oxford: Clarendon Press.
66. Hayne, C. and Free, C. (2014). Hybridized professional groups and institutional work: COSO and the rise of enterprise risk management. *Accounting, Organizations and Society*, 39(5), pp.309–330.
67. Higgins, C., Milne, M.J. and van Gramberg, B. (2015). The Uptake of Sustainability Reporting in Australia. *Journal of Business Ethics*, 129(2), pp.445–468.
68. Higgins, C., Stubbs, W. and Love, T. (2014). Walking the talk(s): Organisational narratives of integrated reporting. *Accounting, Auditing & Accountability Journal*, 27(7), pp.1090–1119.
69. Humphrey, C., O'Dwyer, B. and Unerman, J. (2016). Re-theorizing the configuration of organizational fields: the IIRC and the pursuit of 'Enlightened' corporate reporting. *Accounting and Business Research*, 47(1), pp.30–63.
70. Hwang, H. and Colyvas, J.A. (2010). Problematizing Actors and Institutions in Institutional Work. *Journal of Management Inquiry*, 20(1), pp.62–66.
71. Jabbour, M., Cullen, J. and Crawford, J. (2025). Enterprise risk management: an institutional work perspective. *Accounting Forum*, pp.1–29.
72. Jepperson, R.L. and Meyer, J.W. (1991). Institutions, Institutional Effects, and Institutionalism. In: *Institutional Theory: The Cultural Construction of Organizations, States, and Identities*. Cambridge: Cambridge University Press, pp.37–66.
73. Jørgensen, S., Mjøs, A. and Pedersen, L.J.T. (2021). Sustainability reporting and approaches to materiality: tensions and potential resolutions. *Sustainability Accounting, Management and Policy Journal*, 13(2).
74. Laine, M. (2024). Accounting, reporting and verification of impact: Implications for sustainability: A commentary on Adams et al. (2023) 'Styles of verification and the pursuit of organisational repair: The case of social impact'. *Accounting, Organizations and Society*, 113, pp.101–540.
75. Laine, M., Tregidga, H. and Unerman, J. (2021). *Sustainability accounting and accountability*. 3rd ed. London: Routledge.
76. Lawrence, T., Suddaby, R. and Leca, B. (2011). Institutional Work: Refocusing Institutional Studies of Organization. *Journal of Management Inquiry*, 20(1), pp.52–58.

77. Lawrence, T.B., Leca, B. and Zilber, T.B. (2013). Institutional Work: Current Research, New Directions and Overlooked Issues. *Organization Studies*, 34(8), pp.1023–1033.
78. Lawrence, T.B. and Suddaby, R. (2006). Institutions and Institutional Work. In: S.R. Clegg, C. Hardy, T.B. Lawrence and W.R. Nord, eds., *Sage Handbook of Organization Studies*. London: Sage, pp.215–254.
79. Lawrence, T.B., Suddaby, R. and Leca, B. (2009). *Institutional work : actors and agency in institutional studies of organization*. Cambridge: Cambridge University Press.
80. Litrico, J.-B. and David, R.J. (2017). The Evolution of Issue Interpretation within Organizational Fields: Actor Positions, Framing Trajectories, and Field Settlement. *Academy of Management Journal*, 60(3), pp.986–1015.
81. Lounsbury, M. (2008). Institutional rationality and practice variation: New directions in the institutional analysis of practice. *Accounting, Organizations and Society*, 33(4-5), pp.349–361.
82. Luque-Vílchez, M., Husillos, J. and Larrinaga, C. (2024). The construction of the normative persuasion of social and environmental reporting regulation. *Sustainability Accounting Management and Policy Journal*, 15(7), pp.34–62.
83. Maechler, S. and Graz, J.-C. (2019). The Standardisation of Natural Capital Accounting Methodologies. In: K. Jakobs, ed., *Shaping the Future Through Standardization*. Hershey PA: Engineering Science Reference.
84. Maguire, S. and Hardy, C. (2009). Discourse and Deinstitutionalization: The Decline of DDT. *The Academy of Management Journal*, 52(1), pp.148–178.
85. Merriam, S.B. and Tisdell, E.J. (2016). *Qualitative research: a Guide to Design and Implementation*. 4th ed. San Francisco: Jossey-Bass.
86. Meyer, J.W. and Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology*, 83(2), pp.340–363.
87. Micelotta, E.R. and Washington, M. (2013). Institutions and Maintenance: The Repair Work of Italian Professions. *Organization Studies*, 34(8), pp.1137–1170.
88. Morgan, H. (2022). Conducting a qualitative document analysis. *The Qualitative Report*, 27(1), pp.64–77.

89. O’Dwyer, B. (2004). Qualitative Data Analysis: Illuminating a Process for Transforming a ‘Messy’ but ‘Attractive’ ‘Nuisance’. *The Real Life Guide to Accounting Research*, pp.391–407.
90. O’Dwyer, B. (2011). The Case of Sustainability Assurance: Constructing a New Assurance Service. *Contemporary Accounting Research*, 28(4), pp.1230–1266.
91. O’Dwyer, B. (2024). Constructing global climate-related risk reporting: Organizing dissonance in the Task Force on Climate-related Financial Disclosures (TCFD). Unpublished.
92. O’Dwyer, B., Humphrey, C. and Rowbottom, N. (2024). From institutional integration to institutional demise: The disintegration of the International Integrated Reporting Council (IIRC). *Critical Perspectives on Accounting*, 99, pp.102-699.
93. O’Dwyer, B. and Unerman, J. (2020). Shifting the focus of sustainability accounting from impacts to risks and dependencies: researching the transformative potential of TCFD reporting. *Accounting, Auditing & Accountability Journal*, 33(5), pp.1113–1141.
94. O’Sullivan, N. and O’Dwyer, B. (2015). The structuration of issue-based fields: Social accountability, social movements and the Equator Principles issue-based field. *Accounting, Organizations and Society*, 43, pp.33–55.
95. Oliver, C. (1992). The Antecedents of Deinstitutionalization. *Organization Studies*, 13(4), pp.563–588.
96. Österblom, H., Bebbington, J., Blasiak, R., Sobkowiak, M. and Folke, C. (2022). Transnational Corporations, Biosphere Stewardship, and Sustainable Futures. *Annual Review of Environment and Resources*, 47(1).
97. Pache, A.-C. and Santos, F. (2010). When Worlds Collide: The Internal Dynamics of Organizational Responses to Conflicting Institutional Demands. *Academy of Management Review*, 35(3), pp.455–476.
98. Pelger, C., Sellhorn, T., Wagner, V. and Weiß, K. (2024). Aligning the stars for the global baseline: The problems, policies and politics of creating the ISSB. Unpublished.
99. Pelger, C. and Spieß, N. (2016). On the IASB’s construction of legitimacy – the case of the agenda consultation project. *Accounting and Business Research*, 47(1), pp.64–90.

100. Perkmann, M. and Spicer, A. (2008). How are management fashions institutionalized? The role of institutional work. *Human Relations*, 61(6), pp.811–844.
101. Pfeffer, J. and Salancik, G.R. (1978). *The external control of organizations: A resource dependence perspective*. Stanford, Calif: Stanford Business Books.
102. Powell, W.W. and DiMaggio, P.J. (1991). *The new institutionalism in organizational analysis*. Chicago: University of Chicago Press.
103. Ranzani, P. and Tweedie, J. (2024). On the non-production of an accounting standard: Climate change, carbon emissions, and organisational bullshit. Unpublished.
104. Rainforest Action Network (2023). *Final TNFD Framework Launching September 18th: Rainforest Action Network Response - Rainforest Action Network*. [online] Rainforest Action Network. Available at: https://www.ran.org/press-releases/final-tnfd-framework-launching-september-18th-rainforest-action-network-response/ [Accessed 9 Jun. 2025].
105. Rao, H., Morrill, C. and Zald, M.N. (2000). Power Plays: How Social Movements and Collective Action Create New Organizational Forms. *Research in Organizational Behavior*, 22, pp.237–281.
106. Reuters (2019). AXA says second-half natural disaster claims to exceed usual levels. *Reuters*. [online] 6 Nov. Available at: https://www.reuters.com/article/business/axa-says-second-half-natural-disaster-claims-to-exceed-usual-levels-idUSKBN1XG2HZ/.
107. Rhianon Edgley, C., Jones, M.J. and Solomon, J.F. (2010). Stakeholder inclusivity in social and environmental report assurance. *Accounting, Auditing & Accountability Journal*, 23(4), pp.532–557.
108. Richardson, A.J. and Eberlein, B. (2010). Legitimating Transnational Standard-Setting: The Case of the International Accounting Standards Board. *Journal of Business Ethics*, 98(2), pp.217–245.
109. Rowbottom, N. (2022). Orchestration and consolidation in corporate sustainability reporting. The legacy of the Corporate Reporting Dialogue. *Accounting, Auditing & Accountability Journal*, 36(3).
110. Scott, W.R. (1995). *Institutions and Organizations: ideas, interests, and Identities*. Thousand Oaks: Sage Publications, Inc.

111. See, G. (2025). *'Stop financing harm': TNFD co-chair on how to close the nature finance gap*. [online] Eco-Business. Available at: https://www.eco-business.com/news/stop-financing-harm-tnfd-co-chair-on-how-to-close-the-nature-finance-gap/ [Accessed 9 Jun. 2025].
112. Selznick, P. (1996). Institutionalism 'Old' and 'New'. *Administrative Science Quarterly*, 41(2), p.270.
113. Shapiro, B. and Matson, D. (2008). Strategies of resistance to internal control regulation. *Accounting, Organizations and Society*, 33(2-3), pp.199–228.
114. Slager, R., Gond, J.-P. and Moon, J. (2012). Standardization as Institutional Work: The Regulatory Power of a Responsible Investment Standard. *Organization Studies*, 33(5-6), pp.763–790.
115. Suchman, M.C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *The Academy of Management Review*, 20(3), pp.571–610.
116. Suddaby, R., Cooper, D.J. and Greenwood, R. (2007). Transnational regulation of professional services: Governance dynamics of field level organizational change. *Accounting, Organizations and Society*, 32(4-5), pp.333–362.
117. Suddaby, R. and Greenwood, R. (2005). Rhetorical Strategies of Legitimacy. *Administrative Science Quarterly*, 50(1), pp.35–67.
118. Suddaby, R. and Viale, T. (2011). Professionals and field-level change: Institutional work and the professional project. *Current Sociology*, 59(4), pp.423–442.
119. Tett, G. (2023). *David Craig: Our economic system is completely dependent on nature*. [online] Financial Times. Available at: https://www.ft.com/content/66fa93fa-2766-4eff-8bba-d528fdcc410b [Accessed 12 Mar. 2025].
120. Thomas, L.D.W. and Ritala, P. (2021). Ecosystem Legitimacy Emergence: A Collective Action View. *Journal of Management*, 48(3), pp.515–541.
121. TNFD (2021a). *NATURE IN SCOPE: A summary of the proposed scope, governance, work plan, communication and resourcing plan of the TNFD*.
122. TNFD (2021b). *Q&A with the three co-chairs of the Informal Working Group bringing together a TNFD*. [online] TNFD. Available at:

https://tnfd.global/qa-with-the-three-co-chairs-of-the-informal-working-group-bringing-together-a-tnfd/ [Accessed 12 Mar. 2025].

123. TNFD (2021c). *TNFD Global Launch Event*. [online] YouTube. Available at: https://www.youtube.com/watch?v=b-PqBFfMaJ0 [Accessed 12 Mar. 2025].

124. TNFD (2022a). *The TNFD framework beta v0.1*. [online] Available at: https://tnfd.global/wp-content/uploads/2023/07/TNFD-Full-Report-Mar-2022-Beta-v0-1-1.pdf?v=1690527771 [Accessed 12 Mar. 2025].

125. TNFD (2022b). *The TNFD framework beta v0.2*. [online] *TNFD*. Available at: https://tnfd.global/wp-content/uploads/2023/07/TNFD-Framework-Document-Beta-v0-2-v2-1.pdf?v=1690527784 [Accessed 12 Mar. 2025].

126. TNFD (2023). *Taskforce on Nature-related Financial Disclosures (TNFD) Recommendations*. [online] *TNFD*. Available at: https://tnfd.global/wp-content/uploads/2023/08/Recommendations-of-the-Taskforce-on-Nature-related-Financial-Disclosures.pdf?v=1734112245 [Accessed 12 Mar. 2025].

127. TNFD (2024a). *320 companies and financial institutions to start TNFD nature-related corporate reporting*. [online] TNFD. Available at: https://tnfd.global/320-companies-and-financial-institutions-to-start-tnfd-nature-related-corporate-reporting/?utm_source=chatgpt.com [Accessed 3 Apr. 2025].

128. TNFD (2025). *TNFD – Taskforce on Nature-related Financial Disclosures*. [online] Available at: https://tnfd.global/ [Accessed 11 Mar. 2025].

129. TNFD. (2024b). *Over 500 organisations and $17.7 trillion AUM now committed to TNFD-aligned risk management and corporate reporting*. [online] Available at: https://tnfd.global/over-500-organisations-and-17-7-trillion-aum-now-committed-to-tnfd-aligned-risk-management-and-corporate-reporting/ [Accessed 2 Jan. 2025].

130. Trank, C.Q. and Washington, M. (2009). Maintaining an institution in a contested organizational field: the work of the AACSB and its constituents. In: *Institutional Work: Actors and Agency in Institutional Studies of Organizations*. Cambridge: Cambridge University Press, pp.236–261.

131. UK Government (2021). *Taskforce for Nature-related Financial Disclosures Secretariat: An International Climate Finance Business Case Department of Environment, Food and Rural Affairs*.

132. UNDP (2021). *The UN & TNFD: Stewarding Systems Change*. [online] Available at: https://www.undp.org/news/un-tnfd-stewarding-systems-change [Accessed 11 Mar. 2025].

133. UNDP (2022). *Nature readiness for nature-related disclosures in emerging market: Recommendations from 5 case studies for a successful disclosure framework*. pp.1–52.

134. UNEP (2021). *Nature loss threatens global economy: report*. [online] Available at: https://www.unep.org/news-and-stories/story/nature-loss-threatens-global-economy-report [Accessed 11 Mar. 2025].

135. UNEP FI (2023a). *Insights into UNEP FI's TNFD pilots*. [online] Available at: https://www.unepfi.org/publications/unep-fi-pilots-in-support-of-tnfd/ [Accessed 20 Nov. 2024].

136. UNEP FI (2023b). *Unboxing Nature-related Risks*. [online] Available at: https://www.unepfi.org/wordpress/wp-content/uploads/2023/04/Unboxing-Nature-related-Risks.pdf [Accessed 20 Nov. 2024].

137. Van Burg, E., Cornelissen, J., Stam, W. and Jack, S. (2020). Advancing Qualitative Entrepreneurship Research: Leveraging Methodological Plurality for Achieving Scholarly Impact. *Entrepreneurship Theory and Practice*, 46(1), pp.3–20.

138. Van Wijk, J., Stam, W., Elfring, T., Zietsma, C. and den Hond, F. (2013). Activists and Incumbents Structuring Change: The Interplay of Agency, Culture, and Networks in Field Evolution. *Academy of Management Journal*, 56(2), pp.358–386.

139. Wijen, F. (2014). Means versus Ends in Opaque Institutional Fields: Trading off Compliance and Achievement in Sustainability Standard Adoption. *Academy of Management Review*, 39(3), pp.302–323.

140. World Bank Group (2020). *Mobilizing Private Finance for Nature'*. [online] *World Bank Group*. Available at: https://thedocs.worldbank.org/en/doc/916781601304630850-0120022020/original/FinanceforNature28Sepwebversion.pdf [Accessed 1 Apr. 2025].

141. WWF (2018). *Living Planet Report 2018: Aiming higher*. [online] *WWF*. Available at: https://www.wwf.org.uk/sites/default/files/2018-10/wwfintl_livingplanet_full.pdf [Accessed 9 Jun. 2025].
142. WWF (2019). *WWF in Davos for the World Economic Forum Annual Meeting*. [online] Available at: https://wwf.panda.org/wwf_news/?341715/WWF-in-Davos-for-the-World-Economic-Forum-Annual-Meeting [Accessed 12 Mar. 2025].
143. WWF (2020). *Business leaders and the environmental movement unite in Davos behind a New Deal for Nature and People*. [online] WWF. Available at: https://updates.panda.org/business-leaders-and-the-environmental-movement-unite-in-davos [Accessed 9 Jun. 2025].
144. WWF (2024). *2024 LIVING PLANET REPORT: A System in Peril*. [online] Available at: https://wwflpr.awsassets.panda.org/downloads/2024-living-planet-report-a-system-in-peril.pdf [Accessed 11 Mar. 2025].
145. Yin, R.K. (2018). *Case Study Research and Applications: Design and Methods*. 6th ed. Thousand Oaks: Sage Publications.
146. Young, J.J. (2006). Making up users. *Accounting, Organizations and Society*, 31(6), pp.579–600.
147. Zadek, S. and Chambers, N. (2021). *TNFD Consultation, Technical Scope*. [online] BlackRock. Available at: https://www.blackrock.com/corporate/literature/publication/our-response-to-the-technical-scope-of-the-taskforce-on-nature-related-financial-disclosures-consultation.pdf [Accessed 12 Mar. 2025].
148. Zietsma, C., Groenewegen, P., Logue, D.M. and Hinings, C.R. (Bob) (2017). Field or Fields? Building the Scaffolding for Cumulation of Research on Institutional Fields. *Academy of Management Annals*, 11(1), pp.391–450.
149. Zietsma, C. and Lawrence, T.B. (2010). Institutional Work in the Transformation of an Organizational Field: The Interplay of Boundary Work and Practice Work. *Administrative Science Quarterly*, 55(2), pp.189–221.
150. Zietsma, C. and McKnight, B. (2009). Building the iron cage: institutional creation work in the context of competing proto-institutions. In: *Institutional Work: Actors and Agency in Institutional Studies of Organizations*. Cambridge: Cambridge University Press, pp.143–177.

151. Zilber, T.B. (2002). INSTITUTIONALIZATION AS AN INTERPLAY BETWEEN ACTIONS, MEANINGS, AND ACTORS: THE CASE OF A RAPE CRISIS CENTER IN ISRAEL. *Academy of Management Journal*, 45(1), pp.234–254.

152. Zilber, T.B. (2009). Institutional maintenance as narrative acts. In: T.B. Lawrence, R. Suddaby and B. Leca, eds., *Institutional Work: Actors and Agency in Institutional Studies of Organizations*. Cambridge: Cambridge University Press.

Can Synthetic Data Fine-Tune Financial Language Models? Evidence from Sentiment and Market Reactions

Eghbal Rahimikia[†] Kai Zhang[‡] Ser-Huang Poon[§]

March 27, 2026

Abstract

This paper evaluates whether synthetically generated supervision can replace proprietary labelled data for fine-tuning financial sentiment models without distorting economic inference. Using the same pretrained FinBERT backbone, we compare models fine-tuned on proprietary data and on a consensus-based synthetic corpus. On a manually labelled internal benchmark, the synthetic-trained model achieves higher accuracy and substantially more balanced macro-F_1 performance than the proprietary counterpart, indicating that synthetic supervision provides strong classification signal. In earnings-call regressions, standalone tone measures derived from synthetic supervision produce economically meaningful and statistically significant associations with short-window abnormal returns, with coefficient magnitudes closely aligned to those obtained from the proprietary model.

The two tone measures are highly correlated ($\rho = 0.935$), indicating substantial overlap in captured textual content. However, in joint "horse-race" regressions, coefficient equality is rejected, implying that synthetic and proprietary supervision are not perfectly interchangeable when decomposing marginal signal components. This rejection reflects differential allocation of highly overlapping tone information rather than a reversal of economic conclusions.

Taken together, the results suggest that synthetic supervision is already a viable and transparent alternative for constructing financial tone measures that preserve core pricing content in standalone applications. At the same time, residual differences in marginal loadings indicate scope for further refinement of synthetic labelling protocols. The findings support the use of synthetic data as a scalable substitute where proprietary labels are unavailable, while highlighting the importance of measurement-invariance diagnostics in financial text research.

Keywords: Financial Textual Analysis, Sentiment Measurement, Synthetic Data, Transformer Models in Finance, Earnings Conference Calls

[†] eghbal.rahimikia@manchester.ac.uk, Alliance Manchester Business School, University of Manchester [‡] kai.zhang-13@postgrad.manchester.ac.uk, Alliance Manchester Business School, University of Manchester
[§] ser-huang.poon@manchester.ac.uk, Alliance Manchester Business School, University of Manchester

1 Introduction

Qualitative information conveyed in corporate disclosures plays a central role in financial markets. A substantial literature documents that the tone of earnings conference calls contains incremental information beyond earnings news and is incorporated rapidly into stock prices (Price et al., 2012; Huang et al., 2014b). These findings have motivated the widespread use of textual measures to study managerial communication, investor interpretation, and asset-pricing dynamics.

Over time, the construction of financial sentiment measures has evolved. Early research relies on domain-specific dictionaries tailored to financial language, emphasising that general-purpose linguistic tools may misclassify economic text (Loughran and McDonald, 2011). More recent work increasingly employs supervised machine learning and transformer-based models trained on labelled financial corpora, which can improve contextual understanding and classification performance (Huang et al., 2023; De Kok, 2025). However, an important practical limitation remains: many state-of-the-art financial language models are fine-tuned using proprietary labelled data that are not publicly accessible. This reliance on inaccessible labels raises concerns about transparency, comparability, and replicability in empirical finance.

This paper studies a measurement problem in empirical asset pricing: whether sentiment proxies derived from modern language models remain economically valid when the supervised labels used for fine-tuning are unavailable to other researchers. We use synthetic supervision—labels generated through a structured, multi-model consensus procedure—as a transparent substitute for proprietary training data, and examine whether replacing inaccessible labels with synthetic ones materially alters asset-pricing inference in a standard earnings-call event setting.

We frame synthetic supervision as a measurement-invariance test in asset pricing, grounded in a classical measurement-error framework. Our empirical design treats synthetic supervision as an alternative measurement technology. Identification therefore rests on whether replacing proprietary fine-tuning labels with synthetic labels alters the estimated mapping between textual tone and market reactions in an otherwise fixed economic setting.

Let $S^*_{i,t}$ denote the latent, economically relevant tone embedded in firm i's earnings conference call at time t. The underlying asset-pricing relation can be written as:

$$CAR_{i,t} = \alpha + \beta S^*_{i,t} + \gamma X_{i,t} + \mu_i + \lambda_t + u_{i,t},$$

where $CAR_{i,t}$ denotes short-window cumulative abnormal returns, $X_{i,t}$ is a vector of firm-level controls including earnings news and standard characteristics, and μ_i and λ_t denote firm and time fixed effects, respectively.

In practice, $S^*_{i,t}$ is unobserved. Instead, researchers construct a proxy $S^{(m)}_{i,t}$ using a particular model m, which can be represented as:

$$S^{(m)}_{i,t} = S^*_{i,t} + \varepsilon^{(m)}_{i,t}.$$

If $\varepsilon^{(m)}_{i,t}$ behaves as classical measurement error—mean zero and uncorrelated with $S^*_{i,t}$ and $X_{i,t}$—then replacing $S^*_{i,t}$ with $S^{(m)}_{i,t}$ may attenuate coefficient magnitudes but should not reverse signs or generate spurious significance. By contrast, if the construction of $S^{(m)}_{i,t}$ introduces systematic bias, estimated asset-pricing relations may change in economically meaningful ways.

Our central question is therefore not whether synthetic fine-tuning maximises benchmark classification accuracy, but whether synthetic supervision alters the mapping from text to sentiment in a way that affects economic inference. We operationalise this as a measurement-invariance test: do tone measures extracted from models fine-tuned on synthetic labels yield statistically and economically comparable estimates of β to those obtained from models trained on proprietary labels? Failure to detect economically meaningful differences would be consistent with synthetic supervision acting primarily as additional measurement noise rather than introducing systematic bias.

To investigate this question, we construct consensus-labelled synthetic sentiment data using fixed large language models under a transparent two-stage generation and tagging protocol. We use these data to fine-tune several widely used pretrained encoders, including FinBERT (Huang et al., 2023), BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), and FinText (Rahimikia and Drinkall, 2024).

Our empirical design proceeds in three steps. First, we evaluate classification performance on a manually reviewed synthetic benchmark to ensure internal consistency of the training procedure. Second, we assess external validity using the publicly available Financial PhraseBank dataset (Malo et al., 2013). Third, and most importantly, we construct transcript-level executive tone from earnings conference calls and estimate market-reaction regressions analogous to those in Huang et al. (2023). We then formally test whether estimated tone coefficients differ across models trained on synthetic versus proprietary labels. Throughout, we use conditional language and statistical testing to assess whether any observed differences are economically meaningful.

Our contribution is threefold. First, we address an underexamined measurement issue in financial textual analysis: the dependence of modern language models on proprietary fine-tuning labels. While prior work compares dictionaries and supervised classifiers (Frankel et al., 2022), less attention has been paid to whether the source of training labels itself affects downstream economic conclusions. By treating synthetic supervision as a controlled substitution for inaccessible labels, we provide evidence on whether such substitution preserves or alters asset-pricing inference.

Second, we extend the validation framework for financial text measures beyond benchmark accuracy. Classification metrics are informative about predictive performance relative to labelled datasets, but they do not directly address whether a sentiment proxy retains its economic interpretation. By focusing on the stability of the tone–return relation in an earnings-call setting, we evaluate sentiment measures against an economically meaningful criterion. This approach responds to concerns in the broader NLP literature that benchmark-centric evaluation may not fully capture downstream validity (Galke et al., 2025).

Third, the paper addresses a reproducibility constraint in empirical finance: many state-of-the-art sentiment models rely on proprietary fine-tuning labels that cannot be independently examined. Synthetic supervision provides a transparent substitute. As generative models are increasingly used to construct, label, and transform financial text (Mo and Ouyang, 2025), questions arise about transparency and comparability across studies. If synthetic supervision can be shown to preserve economically relevant inference under clearly documented protocols, it may offer a pathway toward more replicable text-based measures without reliance on proprietary corpora.

Importantly, the scope of our validation is deliberately narrow. We focus on sentiment in earnings conference calls and evaluate economic inference within a standard short-window event-study framework. We do not assume that synthetic data replicate the full distributional properties of real financial language, nor do we claim universal validity across all textual constructs. Rather, we investigate whether, in this specific and economically relevant setting, substituting synthetic supervision for proprietary labels materially changes estimated asset-pricing relations.

Taken together, the analysis is designed to clarify whether synthetic supervision behaves like additional measurement noise or introduces systematic bias in financial sentiment proxies. The answer to this question has direct implications for transparency, replicability, and construct validity in empirical asset pricing.

2 Synthetic Data Construction and Fine-Tuning Framework

The synthetic-data framework is designed with a narrowly defined objective. Synthetic text is used exclusively to generate labelled training data for fine-tuning encoder-based classifiers when proprietary labelled financial corpora are unavailable. Synthetic data are not used to simulate economic outcomes or construct artificial financial variables. Rather, they provide a transparent labelling mechanism whose validity is evaluated through (i) classification performance (Table 1) and (ii) downstream economic inference in earnings-call regressions (Section 4).

Figure 1 illustrates the complete pipeline: sentence generation, post-processing, deterministic multi-model tagging, consensus filtering, corpus construction, fine-tuning, and multi-layer evaluation. The framework is evaluated primarily for sentiment classification, where economically interpretable validation is feasible. We do not claim general validity for all linguistic constructs.

2.1 Two-Stage Generation and Tagging Protocol

We adopt a two-stage procedure consisting of (i) high-temperature sentence generation and (ii) deterministic cross-model tagging. Full prompts and task-specific definitions are reported in Appendix A.

2.1.1 Stage 1: Finance-Domain Sentence Generation

Synthetic sentences are generated using four frontier large language models: GPT-4, Gemini 1.5, Mistral Large, and Claude Opus. Generation temperature is set at the model default (approximately $T \approx 1$) to encourage lexical diversity while maintaining finance-domain realism. Prompts emulate earnings conference calls, analyst reports, regulatory filings, and related financial disclosures. For the sentiment task, each generation call produces a batch of 15 sentences: 5 positive, 5 neutral, and 5 negative. Prompts explicitly prohibit identifiable entities (e.g., firm names, dates, events) to prevent anchoring to specific firms or time periods. Between generation and tagging, we apply systematic post-processing: removal of empty lines and introductory phrases, regular-expression cleaning of annotations, splitting numbered outputs into individual sentences, random shuffling to mitigate order effects, and formatting standardisation. These steps are applied uniformly across generators.

2.1.2 Stage 2: Deterministic Multi-Model Tagging

After generation, we implement a structured tagging protocol designed to reduce single-model bias. When a given model serves as the generator, the remaining models independently assign sentiment labels. Tagging is conducted deterministically ($T = 0$) to eliminate stochastic variation. Each tagger must assign one of the predefined labels (*Positive*, *Neutral*, *Negative*) or return `None` if no confident classification is possible. Each sentence is then classified into one of three consensus outcomes:

- **Unanimous:** all taggers assign the same non-`None` label,
- **2-vs-1:** two taggers agree and one disagrees,
- **All-different:** no pair agrees.

Panel A of Table 1 reports the resulting consensus distributions by generator. Unanimous agreement rates are high across generators: 91.83% (GPT), 88.89% (Gemini), 88.69% (Mistral), and 91.18% (Opus). Two-versus-one rates range from 8.17% to 11.30%, while all-different outcomes are negligible (at most 0.06%). These statistics provide transparency regarding cross-model label formation. Importantly, consensus rates are descriptive. They do not imply predictive superiority; downstream performance must be assessed empirically.

2.2 Construction of Synthetic Sentiment Corpora

We construct two distinct corpora for sentiment: (i) a consensus-based fine-tuning corpus and (ii) a human-labelled benchmark corpus.

2.2.1 Consensus-Based Fine-Tuning Corpus

The fine-tuning corpus is constructed exclusively from sentences that receive unanimous non-`None` labels under the multi-model tagging protocol described in Section 2.1. Panel A of Table 1 reports the full inventory of unanimously labelled sentences by generator model. These inventories are substantial (e.g., 2,840 GPT-positive, 5,884 Gemini-positive, 3,149 Mistral-positive, and 5,389 Opus-positive), and class proportions differ across generators, reflecting stylistic variation in generation rather than imposed balance.

For fine-tuning, training data are not restricted to any single generator. Instead, unanimously labelled sentences are pooled across all generators into a combined inventory and partitioned by sentiment class (Positive, Neutral, Negative). From each class-specific pool, 500 observations are drawn using uniform random sampling without replacement, yielding a balanced training set of 1,500 sentences (500 per class).[1]

Random subsampling serves three methodological purposes. First, it imposes exact class balance despite generator-level imbalances documented in Panel A. Second, fixing the training size isolates architectural and representation effects from sample-size effects. Third, pooling across generators mitigates reliance on any single model's stylistic footprint, thereby reducing generator-specific bias.

The unanimous-only filter prioritises label precision over sample size, consistent with a bias–variance trade-off in measurement construction. All main experiments are replicated across 10 random seeds; these replications vary optimisation initialisation but not the underlying training corpus.

[1]The 1,500-sentence training set (500 per class) is drawn once from the pooled unanimous inventory and held fixed across all encoder architectures to ensure that performance differences reflect model characteristics rather than variation in training samples. Random seed replication affects optimisation initialisation only, not the composition of the training corpus.

2.2.2 Human-Labelled Synthetic Benchmark Corpus

To evaluate performance beyond consensus-easy cases, we construct a separate benchmark corpus retaining both:

- **Matched** sentences (two taggers agree),
- **Unmatched** sentences (two taggers disagree).

Each benchmark sentence is manually reviewed, and the manual label is treated as the ground truth.

Panel B of Table 1 reports benchmark composition and agreement rates relative to the manual benchmark. The matched subset contains 60 sentences (22 positive, 15 neutral, 23 negative), with pooled agreement of 88%. The unmatched subset contains 45 sentences (13 positive, 11 neutral, 21 negative), with pooled agreement of 49%. This sharp difference confirms that disagreement identifies semantically ambiguous observations. The benchmark therefore enables controlled comparison of classifier performance across high-certainty (matched) and low-certainty (unmatched) observations.

2.3 Encoder Architectures and Fine-Tuning Protocol

We fine-tune four widely used pretrained encoder backbones—FinBERT, BERT, RoBERTa, and FinText (Base and Small)—to ensure that our conclusions do not depend on a single model family. These architectures differ in pretraining corpus, domain adaptation, and parameterisation. Consistent results across backbones therefore mitigate concerns that findings are driven by representation-specific artefacts rather than the supervision scheme itself.

To enable meaningful cross-model comparison, we impose a fully standardised fine-tuning protocol. All models are trained on the same fixed synthetic training corpus described in Section 2.2.1, consisting of 1,500 sentences (500 per sentiment class). Training size is therefore held constant across architectures. Optimisation parameters are likewise standardised. We use a learning rate of 2×10^{-5}, a batch size of 32, and replicate each configuration across 10 random seeds. Seed variation affects only optimisation initialisation, not the composition of the training data. This design isolates architectural differences from sample-driven variation. Within this otherwise fixed framework, we conduct a targeted hyperparameter search over training epochs. Each model is estimated under four candidate horizons: $\{3, 5, 7, 10\}$ epochs. Performance is evaluated on two held-out datasets: (i) the human-labelled synthetic benchmark corpus (Section 2.2.2), and (ii) the external Financial PhraseBank dataset. This dual evaluation prevents selection of an epoch that overfits idiosyncrasies of a single test environment and provides an out-of-domain check on generalisation.

Figures 2–3 summarise the grid-search results. Across architectures, performance improves materially from 3 to 5 epochs but exhibits diminishing returns thereafter. Gains at 7 or 10 epochs are small relative to cross-seed variation. We therefore adopt five epochs as a common training horizon for all subsequent experiments. By fixing training data, optimiser settings, and epoch choice ex ante under a transparent grid search, we minimise hyperparameter-induced comparability concerns and ensure that performance differences reflect architectural and supervision effects rather than tuning discretion.

2.4 Scope and Cross-Task Generalisation of the Synthetic Pipeline

Although sentiment serves as the primary illustration, the same generation–tagging–fine-tuning protocol is applied to additional financial text tasks, including ESG classification (4- and 11-category), emotion categorisation, forward-looking statement detection, and overconfidence identification. Online Appendix Tables 10 and 11 indicate that the consensus mechanism scales procedurally beyond sentiment.

Across tasks, the consensus mechanism exhibits consistently high cross-model agreement. Panel A of Table 10 shows that unanimous rates are uniformly elevated across categories. For ESG4, unanimous agreement ranges from 84.21% (Opus) to 95.87% (GPT). For ESG11, rates remain between 87.59% and 95.94%. Emotion classification records unanimous rates between 85.55% and 96.46%, while overconfidence reaches as high as 97.52%. Even the forward-looking statement (FLS) task—where conceptual boundaries are inherently more nuanced—achieves unanimous rates between 67.82% and 93.17%. All-different outcomes are economically negligible across tasks, generally below 1% and never exceeding 2.78%. These figures indicate that the deterministic multi-model tagging protocol scales procedurally beyond sentiment and produces internally consistent labels across heterogeneous financial constructs.

Human-benchmark validation in Table 11 confirms that consensus intensity is economically meaningful. For ESG4, pooled agreement in the matched subset is 91.2%, compared with 44.4% in the unmatched subset. For ESG11, matched agreement reaches 92.7% versus 41.8% for unmatched observations. Emotion tasks display a similar pattern (85.6% matched versus 35.8% unmatched), as do FLS (91.7% versus 44.8%) and overconfidence (90.0% versus 52.6%). The large matched–unmatched gap across all tasks indicates that cross-model consensus effectively identifies lower-ambiguity sentences, while disagreement captures semantically difficult cases.

Although procedural agreement is high, classification performance varies systematically with task complexity. Structured taxonomies such as ESG4 and ESG11 exhibit relatively stable performance dispersion across architectures, consistent with their clearer categorical boundaries. By contrast, higher-dimensional or behaviourally nuanced constructs (e.g., emotion and overconfidence) show lower benchmark accuracy and greater variation across models, reflecting greater semantic overlap and subjectivity. The FLS task occupies an intermediate position: while consensus rates are high, the wider dispersion in unanimous rates across generators (67.82%–93.17%) suggests that forward-looking language involves greater interpretive variability at the generation stage.

Taken together, Tables 10 and 11 reveal two central patterns. First, the synthetic generation–consensus protocol scales consistently across heterogeneous financial text domains, delivering high internal agreement and strong validation within the matched human-benchmark subsets. Second, predictive performance remains construct-specific. Procedural consistency in label formation does not imply uniform classification accuracy or economic validity across tasks, underscoring the need for construct-level evaluation.

Overall, the synthetic pipeline provides a fully documented and replicable alternative to proprietary fine-tuning. The next section therefore examines whether substituting this transparent supervision mechanism preserves the estimated tone–return relation under formal measurement-invariance tests.

3 Performance of Fine-Tuned Encoders on Sentiment Classification

We evaluate sentiment classification along two dimensions: (i) internal validation on a manually reviewed synthetic benchmark, and (ii) external validation on the Financial PhraseBank dataset. Performance is summarised using macro-averaged Accuracy, Precision, Recall, and F_1 statistics reported in Panel C of Table 1.

3.1 Internal Validation: Synthetic Benchmark

Panel C(a) of Table 1 evaluates performance on the human-labelled synthetic benchmark. A central comparison concerns FinBERT and FinBERT$_{\text{syn}}$, which share the same pretrained backbone but differ in fine-tuning data. FinBERT is fine-tuned on proprietary labelled data, whereas FinBERT$_{\text{syn}}$ is fine-tuned exclusively on the consensus-labelled synthetic corpus. Because the pretrained architecture is identical, performance differences isolate the effect of the fine-tuning dataset.

On the synthetic benchmark, FinBERT achieves 72.38% accuracy and 62.00 macro-F_1. FinBERT$_{\text{syn}}$ improves to 75.71% accuracy and 74.92 macro-F_1. Recall becomes markedly more balanced. The proprietary FinBERT model records recall of 82.86% (Positive), 15.38% (Neutral), and 97.73% (Negative), indicating strong asymmetry. In contrast, FinBERT$_{\text{syn}}$ reports 67.43%, 80.77%, and 79.32% recall across the three classes, respectively, reflecting substantially more even class performance.

The remaining synthetic models (BERT$_{\text{syn}}$, RoBERTa$_{\text{syn}}$, FinTextBase$_{\text{syn}}$, and FinTextSmall$_{\text{syn}}$) are all fine-tuned using the same synthetic training set. Differences among them therefore reflect backbone architecture rather than variation in supervision. On this benchmark, FinTextBase$_{\text{syn}}$ achieves the highest accuracy (79.62%) and macro-F_1 (78.44), followed by RoBERTa$_{\text{syn}}$ (78.86% accuracy), and BERT$_{\text{syn}}$ (76.95%). FinTextSmall$_{\text{syn}}$ records 74.00% accuracy. These results indicate that, holding the synthetic fine-tuning corpus fixed, model capacity and pretraining design continue to influence performance.

3.2 External Validation: Financial PhraseBank

Panel C(b) of Table 1 reports performance on the Financial PhraseBank dataset. FinBERT achieves 91.70% accuracy and 89.71 macro-F_1, substantially higher than all synthetic fine-tuned models. FinBERT$_{\text{syn}}$ records 71.20% accuracy, RoBERTa$_{\text{syn}}$ 76.94%, FinTextBase$_{\text{syn}}$ 67.77%, FinTextSmall$_{\text{syn}}$ 66.77%, and BERT$_{\text{syn}}$ 62.19%.

The FinBERT–FinBERT$_{\text{syn}}$ comparison is again informative. Because both models share the same pretrained backbone, the performance gap on PhraseBank reflects differences in fine-tuning supervision. However, if the proprietary training data used for FinBERT overlap with or resemble the PhraseBank corpus, superior performance may partly reflect domain alignment rather than intrinsic measurement quality. By contrast, synthetic fine-tuning is entirely independent of the PhraseBank dataset, making its evaluation less susceptible to data-alignment advantages.

It is also important to distinguish two dimensions of comparison. Comparing FinBERT with the synthetic models combines model heterogeneity (different backbones) with differences in fine-tuning data. In contrast, comparisons among FinBERT$_{\text{syn}}$, BERT$_{\text{syn}}$, RoBERTa$_{\text{syn}}$, FinTextBase$_{\text{syn}}$, and FinTextSmall$_{\text{syn}}$ hold the synthetic training set fixed and therefore isolate architectural strength. Panel C shows that backbone choice materially affects cross-domain generalisation, with $\text{RoBERTa}_{\text{syn}}$ (76.94% accuracy) outperforming $\text{FinBERT}_{\text{syn}}$ (71.20%), BERT_{syn} (62.19%), $\text{FinTextBase}_{\text{syn}}$ (67.77%) and $\text{FinTextSmall}_{\text{syn}}$ (66.77%).

Taken together, Panel C suggests that synthetic supervision provides competitive in-domain performance and yields balanced class behaviour. External benchmark dominance by FinBERT should be interpreted cautiously, as it may reflect training-data alignment. The economically relevant question, addressed in Section 4, is whether these classification differences translate into materially different tone–stock return relations.

3.3 Performance on Other Financial Text Classifications

Table 12 of the online Appendix reports internal synthetic-benchmark performance for ESG4, ESG11, Emotions, Forward-looking statements (FLS), and Overconfidence. Two patterns emerge.

First, structured ESG classifications deliver the strongest and most stable results. For ESG4, accuracy ranges from 77.13% ($\text{FinBERT}_{\text{syn}}$) to 78.56% ($\text{RoBERTa}_{\text{syn}}$), with macro-$F_1$ between 73.48 and 74.95, all exceeding the proprietary FinBERT baseline (74.38% accuracy; 71.23 F_1). When the taxonomy expands to ESG11, accuracy compresses into a narrow band (73.13%–73.81%) with F_1 between 72.14 and 72.94, indicating reduced dispersion across architectures as category granularity increases.

Second, more behaviourally nuanced tasks exhibit lower and more dispersed performance. Emotion classification achieves 60.50%–66.77% accuracy (F_1: 60.84–66.89), while Overconfidence ranges from 59.17% to 66.00% (F_1: 58.66–65.72). In contrast, the FLS task shows a marked improvement under synthetic fine-tuning: FinBERT records 51.43% accuracy and 48.19 F_1, whereas synthetic models cluster tightly around 74.48%–76.95% accuracy and 75.33–77.61 F_1.

Overall, holding the synthetic supervision fixed, performance differences primarily reflect backbone architecture, while attainable accuracy varies systematically with task complexity. Structured ESG tasks achieve the highest and most stable scores, whereas affective and behavioural classifications remain more demanding.

4 Market Reaction to Earnings Call Sentiment: A Measurement-Invariance Test

This section implements the measurement-invariance framework outlined in the Introduction. Our objective is not to compare classification accuracy per se, but to assess whether replacing proprietary supervision with synthetic supervision materially alters asset-pricing inference in a standard earnings-call setting.

Let $Tone_{i,t}^{(m)}$ denote transcript-level executive sentiment constructed using model m. If synthetic supervision introduces only classical measurement error relative to a benchmark model, then differences in estimated tone coefficients should be limited to attenuation or sampling variation. By contrast, systematic bias in the construction of $Tone_{i,t}^{(m)}$ could lead to economically meaningful differences in the estimated tone–CAR relation.

Empirically, the implementation proceeds in three steps. First, we describe the data and variable construction. Second, we estimate baseline market-reaction regressions relating tone to short-window abnormal returns. Third, we formally test coefficient equality between benchmark and synthetic sentiment measures.

4.1 Baseline Regression Specification

To assess whether economic inference is invariant to how executive sentiment is measured, we estimate short-window market-reaction regressions in the spirit of Price et al. (2012). Our baseline specification is

$$CAR_{i,t} = \alpha + \beta\, Tone_{i,t} + \gamma' X_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t}, \tag{1}$$

where $CAR_{i,t}$ is the three-day cumulative abnormal return around firm i's earnings conference call in quarter t. $Tone_{i,t}$ is transcript-level executive sentiment constructed from executive speech in the call transcript using a given sentiment model, $X_{i,t}$ includes earnings news and a standard set of firm-level controls, μ_i denotes firm fixed effects, and λ_t denotes year-quarter fixed effects. Standard errors are clustered at the firm level.

To facilitate economically meaningful comparisons across sentiment constructions, we standardise each tone measure prior to estimation (mean zero, unit variance). Coefficients can therefore be interpreted as the change in CAR associated with a one-standard-deviation increase in executive sentiment. Under a measurement-error perspective, invariance would be reflected in stable signs and economically similar magnitudes of $\hat{\beta}$ across models after scale normalisation; we later complement these comparisons with formal cross-model coefficient-equality tests.

Benchmark versus synthetic-trained sentiment models. Our benchmark tone measure is based on FinBERT fine-tuned for sentiment classification following the procedure in Huang et al. (2023) on a researcher-labelled sample of financial analyst report sentences originally compiled by Huang et al. (2014a). This benchmark training set contains 10,000 sentences (3,577 positive; 4,586 neutral; 1,837 negative). We compare this benchmark against alternative tone measures constructed from models fine-tuned on our synthetic data, which are designed to provide a transparent and reproducible substitute for proprietary or inaccessible labels.

Baseline evidence from market reactions. Table 4 reports the baseline estimates for six alternative tone measures under identical controls and fixed effects. Across all specifications, executive tone loads positively and is highly statistically significant. Using the benchmark FinBERT model, the estimated loading is $\hat{\beta} = 1.536$ ($t = 22.58$), implying that a one-standard-deviation increase in executive sentiment is associated with approximately 1.5 percentage points higher three-day abnormal returns around the call. Importantly, the synthetic-data models yield closely aligned estimates: coefficients range from 1.312 (FinTextSmall) to 1.496 (FinTextBase), and remain strongly significant throughout (t-statistics between 18.37 and 21.40). Relative to the benchmark, the dispersion in magnitudes is modest and does not alter the qualitative inference that more positive executive sentiment predicts a stronger short-window market response. Taken together, Table 4 provides initial evidence that, in this baseline setting, sentiment models fine-tuned on synthetic data can preserve the downstream asset-pricing conclusions obtained with a benchmark model trained on researcher-labeled financial text.

4.2 Formal Tests of Coefficient Equality

To assess directly whether synthetic fine-tuning alters economic inference, we conduct formal coefficient-equality tests between alternative tone measures. For a given pair of models—a designated benchmark B and an alternative A—we estimate a two-tone "horse-race" specification in which both tone measures enter jointly:

$$CAR_{i,t} = \alpha + \beta_B \, Tone^{(B)}_{i,t} + \beta_A \, Tone^{(A)}_{i,t} + \gamma' X_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t}, \tag{2}$$

and test

$$H_0: \; \beta_B = \beta_A. \tag{3}$$

All tone variables are standardised within the estimation sample, so (β_B, β_A) capture the incremental association between a one-standard-deviation change in each tone measure and three-day abnormal returns, *conditional on the other tone proxy.* Standard errors are clustered by firm, and all specifications include firm and year-quarter fixed effects.

Table 6 reports the pairwise horse-race estimates and the Wald p-values for (3). The horse-race design is intentionally stringent: when two sentiment proxies are highly correlated, the regression mechanically allocates their *shared* signal across (β_B, β_A), so the estimates should be interpreted as relative "signal ownership" rather than as a replication of the single-tone estimates in Table 4. Accordingly, rejection of (3) indicates that two measures are not interchangeable *within the same regression* (i.e., they load differently on the shared component), even if they yield similar standalone tone–return relations when used separately.

Panel A benchmarks the proprietary-labelled FinBERT tone against synthetic-trained alternatives. In every pairing, the benchmark FinBERT retains a large, positive, and precisely estimated coefficient ($\hat{\beta}_B$ ranges from 1.751 to 2.403 with t-statistics between 12.39 and 18.47), whereas the alternative proxy contributes little incremental explanatory power once FinBERT is included. For example, when FinBERT and FinBERT_syn enter jointly, $\hat{\beta}_{\text{FinBERT}} = 2.055$ ($t = 13.46$) while $\hat{\beta}_{\text{FinBERT\_syn}} = -0.317$ ($t = -2.11$), and the Wald test strongly rejects equality ($p < 0.001$). A similar pattern holds when FinBERT is paired with RoBERTa or FinTextBase, where the alternative coefficients are close to zero and statistically weak, while equality is rejected. These results suggest that, in joint regressions, the benchmark tone measure subsumes much of the return-relevant variation captured by other proxies.

Panels B–E compare synthetic-trained measures against one another and reveal more nuanced patterns. In Panel B, using FinBERT_syn as the benchmark, both FinBERT_syn and RoBERTa load positively when entered jointly ($\hat{\beta}_{\text{FinBERT\_syn}} = 0.712$, $t = 3.65$; $\hat{\beta}_{\text{RoBERTa}} = 1.047$, $t = 5.43$), and the equality restriction cannot be rejected ($p = 0.378$), consistent with these two synthetic-trained tone measures containing comparably return-relevant components in this pairwise contest. Panel D yields an analogous non-rejection when RoBERTa is benchmarked against FinTextBase ($p = 0.332$), again indicating close parity for that pair. In contrast, several horse races involving FinTextSmall exhibit negative alternative loadings once the benchmark proxy is included (e.g., Panel E: $\hat{\beta}_{\text{FinTextBase}} = 2.934$, $t = 14.86$ versus $\hat{\beta}_{\text{FinTextSmall}} = -1.327$, $t = -6.94$; $p < 0.001$), suggesting that FinTextSmall captures residual variation that is differently aligned with short-window returns after conditioning on other tone measures.

Overall, Table 4 provides a stringent test of measurement invariance by estimating pairwise "horse-race" regressions in which two tone proxies enter jointly and evaluating Wald tests of coefficient equality. Across specifications, the benchmark FinBERT proxy typically absorbs the priced component of tone, while alternative proxies contribute little incremental explanatory power once FinBERT is included. In contrast, comparisons among synthetic-trained proxies reveal substantial commonality: several RoBERTa-family measures trained on the same synthetic labels exhibit statistically indistinguishable loadings, whereas other pairings—most notably those involving the smaller-capacity FinText model—display sharp differences. Taken together with Table 4, these findings suggest that synthetic supervision preserves much of the economic signal captured by proprietary labels in standalone asset-pricing regressions, while joint specifications highlight differences in how competing proxies allocate that shared signal when entered simultaneously, consistent with heterogeneous measurement-error structure rather than wholesale changes in economic inference.

4.3 Pairwise Horse-Race Regressions and Coefficient Equality

Table 6 reports pairwise "horse-race" regressions in which two standardised tone measures enter jointly, alongside the full set of controls, firm fixed effects, and year–quarter fixed effects. Each row compares a benchmark tone measure with an alternative and reports (i) the two slope coefficients, (ii) their t-statistics, and (iii) the p-value from a Wald test of coefficient equality $H_0 : \beta_{\text{Bench}} = \beta_{\text{Alt}}$. All specifications are estimated on the same sample ($N = 27{,}291$; adjusted R^2 between 0.050 and 0.064), with standard errors clustered by firm.

Panel A: Benchmark = FinBERT. When FinBERT is paired with alternative models, its coefficient remains economically large and strongly significant, ranging from 1.751 ($t = 15.27$) against BERT to 2.403 ($t = 18.47$) against FinTextSmall. The alternative coefficients are generally small in magnitude (e.g., 0.050 for BERT; -0.027 for FinTextBase) and often statistically weak. In all five comparisons, the Wald test rejects coefficient equality ($p = 0.000$ in every case), indicating that FinBERT captures a distinct component of the tone–return relation relative to the alternatives in joint estimation.

Panel B: Benchmark = FinBERT_syn. Using FinBERT_syn as the benchmark, its coefficient ranges from 0.391 ($t = 1.73$; vs. FinTextBase) to 2.849 ($t = 14.90$; vs. FinTextSmall). Against RoBERTa, the Wald test does not reject equality ($p = 0.378$), suggesting statistically similar pricing slopes in that pairing. However, equality is rejected when paired with BERT ($p = 0.000$), FinTextBase ($p = 0.030$), and FinTextSmall ($p = 0.000$). These results indicate that, in joint regressions, FinBERT_syn and RoBERTa exhibit the closest slope alignment.

Panel C: Benchmark = BERT. When BERT is benchmarked, its coefficient ranges from 0.227 ($t = 1.51$; vs. FinTextBase) to 1.344 ($t = 9.65$; vs. FinTextSmall). In all three pairings, equality is rejected ($p = 0.000$), implying statistically distinct contributions relative to RoBERTa, FinTextBase, and FinTextSmall.

Panel D: Benchmark = RoBERTa. RoBERTa's coefficient equals 0.697 ($t = 3.56$) against FinTextBase and 2.560 ($t = 15.15$) against FinTextSmall. Coefficient equality is not rejected versus FinTextBase ($p = 0.332$), but is rejected versus FinTextSmall ($p = 0.000$). This pattern mirrors Panel B, where RoBERTa and FinBERT_syn also exhibited statistically similar slopes.

Panel E: Benchmark = FinTextBase. When compared with FinTextSmall, FinTextBase loads at 2.934 ($t = 14.86$), while FinTextSmall enters at -1.327 ($t = -6.94$).

Equality is rejected ($p = 0.000$), indicating economically and statistically distinct joint effects.

Overall Patterns. Three regularities emerge. First, equality is rejected in the majority of pairwise comparisons, indicating that when two tone measures enter jointly, their estimated slopes often differ statistically. Second, RoBERTa and FinBERT_syn display the closest alignment, with non-rejection of equality in Panels B and D ($p = 0.378$ and $p = 0.332$, respectively). Third, comparisons involving FinTextSmall frequently show large and oppositely signed alternative coefficients (e.g., -0.795, -1.245, -0.952, -1.327), accompanied by consistent rejection of equality.

Taken together, Table 6 indicates that although tone measures are constructed from related textual signals, their joint pricing loadings are often statistically distinct. Instances of non-rejection (notably between RoBERTa and FinBERT_syn, and between RoBERTa and FinTextBase) suggest partial overlap in pricing content, whereas widespread rejection elsewhere points to model-specific variation in the measured tone signal.

4.4 Robustness and Stability Analyses

The following analyses evaluate whether any observed invariance (or divergence) is sensitive to modelling choices or sample composition.

4.4.1 Alternative Event Windows

We re-estimate the baseline specification using alternative short-window return intervals (e.g., $[-1, +1]$, $[0, +2]$, $[-2, +2]$). If the tone–CAR relation reflects a stable economic mechanism rather than window-specific noise, estimated coefficients should exhibit similar sign and economic magnitude across windows. Substantial variation across windows would suggest that the interpretation of tone depends on event specification.

4.4.2 Alternative Abnormal Return Specifications

To examine sensitivity to the construction of abnormal returns, we estimate alternative return adjustments, including market-adjusted returns, Fama–French three-factor residuals, and Carhart four-factor residuals.

Under measurement invariance, the qualitative relation between tone and abnormal returns should not depend materially on the asset-pricing benchmark. If coefficient estimates vary substantially in sign, magnitude, or statistical inference across return constructions, this would indicate that economic interpretation of tone is sensitive to the chosen adjustment model.

4.4.3 Subsample Stability

Table 7 examines whether the tone–CAR relation varies across firm environments in a way that would undermine our central claim that synthetic supervision can substitute for proprietary labels without materially altering asset-pricing inference. We estimate pooled interaction specifications of $CAR[-1, +1]$ on a standardised tone measure, including firm and year–quarter fixed effects, with standard errors clustered by firm. For each partition,

β denotes the tone loading in the baseline group (Small firms / Low coverage / Pre-2020), while Δ captures the incremental loading for the indicated group (Big / HighCov / Post2020); reported p-values are Wald tests of $H_0 : \Delta = 0$.

Across all tone measures, β is positive and precisely estimated in every baseline group, indicating a robust positive association between tone and short-window announcement returns. The only economically meaningful heterogeneity appears along firm size: Panel A shows uniformly negative and highly significant Δ estimates, implying that the tone–CAR relation is stronger among smaller firms and attenuates for larger firms. By contrast, Panel B provides limited evidence of differential effects by analyst coverage: Δ estimates are small and statistically indistinguishable from zero across specifications, suggesting that the tone channel is not materially conditioned on information-intermediary intensity. Panel C yields modest attenuation post-2020 (negative Δ), with at most marginal significance in a subset of models, consistent with mild time variation but no qualitative break.

Importantly for our measurement argument, the sign patterns and inference on Δ are highly consistent across proprietary-trained and synthetic-trained tone measures, implying that any contextual heterogeneity is shared and does not depend on the origin of supervision. Overall, the interaction results support subsample stability in the sense relevant to our paper: synthetic supervision preserves the same economic signal and delivers comparable conclusions about when (and for whom) tone matters.

5 Market Reaction to Dow Jones News

5.1 Data and Preprocessing

Our textual data are drawn from the *Dow Jones Newswires Text News Feed* from January 2013 to December 2023. The raw feed contains 20,761,104 records (`NewsRaw`). Because the raw newswire includes repeated transmissions, fragmented updates, and records that do not correspond cleanly to a single firm-level information event, we apply a sequence of preprocessing steps to construct a sample suitable for asset-pricing tests.

we construct a `NewsChained` dataset by consolidating multiple documents that correspond to the same underlying news item. Specifically, news sharing the same internal article identifier and timestamp are merged, and the other news with the same identifier and the earliest timestamp are kept. This chaining procedure reduces the sample to 14,599,698 observations, which move from raw feed records to economically meaningful news items

Following Chen et al. (2022), we next restrict attention to news items that are related to exactly one stock and remove news with fewer than 100 characters or exceeding 100,000 characters, yielding 1,399,262 observations (`NewsFiltered`). Focusing on news with a unique stock, ensuring each news pertains to one firm. Therefore, it improves the precision of the mapping from text to the cross-section of stock returns.

We also clean the article text to remove formatting artifacts and boilerplate content[2] The objective is to retain the economically relevant narrative conveyed to investors, while excluding mechanical or publisher-generated content that is unlikely to carry incremental information about firm fundamentals, cash flows, or discount rates.

[2]Include XML markup, encoded HTML symbols, URLs, email addresses, phone numbers, source or contact lines, and repeated non-informational lines, while preserving the main paragraph text of the article body.

We then merge NewsFiltered with stock-level market data and retain only observations with valid one-day open-to-open returns and market capitalization data, resulting in 262,980 observations (Newswith1Stock). This step ensures that each news item can be matched to a corresponding return realization and firm size measure, which is necessary for portfolio formation and return-based inference.

Finally, to reduce redundancy in the news sample, we follow a document-level de-duplication procedure based on bag-of-words representations. Specifically, for each article, we compute its cosine similarity relative to articles released in the prior five business days and exclude the article if its cosine similarity is 0.8 or higher with any earlier article in that window. The resulting de-duplicated sample, denoted NewsDeduplicate, contains 1,182,441 observations. Within this final sample, 20,430 news items contain only a headline, while 1,162,011 contain both a headline and body text. In constructing the tone measure, we combine the headline and body text whenever both are available. The construction of the tone variable is described in Section 5.2.

Table 8 reports the sample construction procedure and the number of observations remaining after each filtering step. Figures 4a–4c provide additional descriptive evidence on the time-series distribution of the final news sample. Figure 4a shows that the annual number of news items is relatively stable over the sample period, with somewhat higher coverage in the later years. Figure 4b shows moderate seasonality across calendar months, although the volume of news remains broadly dispersed throughout the year. Figure 4c shows pronounced intraday clustering in news releases, with clear peaks around market opening hours and in the late afternoon.

5.2 Tone of News

We measure the tone of each news item using sentence-level classification models. Specifically, for each article in NewsDeduplicate, we split the available text into sentences and apply five sentiment classifiers: FinBERT, FinBERT_syn, BERT, RoBERTa, FinTextBase, and FinTextSmall. Each model assigns a sentiment label to every sentence.

For a given model, we define the tone of a news item as

$$\text{Tone}_i = \frac{N_i^+ - N_i^-}{N_i}, \tag{4}$$

where N_i^+ is the number of sentences classified as positive for news item i, N_i^- is the number of sentences classified as negative, and N_i is the total number of sentences in that news item. This construction yields a normalized tone measure bounded between -1 and 1, with larger values indicating more positive news content.

For news items that contain only a headline, the tone measure is based solely on the headline. For news items that contain both a headline and body text, we concatenate the headline and body text and compute tone from the combined text. Thus, our baseline article-level measure is intended to capture the overall tone conveyed to investors by the full textual content of the news item, rather than the headline or body in isolation.

Figure 5 reports the average yearly number of positive and negative news articles, together with the average yearly tone, across the six sentiment classifiers. For each model, we classify each article according to its sentiment label and then compute the yearly counts of positive and negative articles as well as the yearly average of the article-level tone measure. The figure indicates that all classifiers produce more positive than negative articles on average, although the relative balance between positive and negative classifications varies across models. The average yearly tone also differs meaningfully across classifiers, suggesting that alternative sentiment models, including models trained with synthetic supervision, generate systematically different article-level tone signals. Such variation is potentially important for the empirical analyses that relate news tone to subsequent stock returns.

5.3 Portfolios Performance

We implement a zero-net-investment news-tone trading strategy based on the cross-sectional ranking of firm-level tone signals. On each trading day t, we sort stocks by their news tone and form a long–short portfolio that buys the top 20% of stocks with the most positive tone and shorts the bottom 20% of stocks with the most negative tone.

To construct the information set used for portfolio formation on trading day t, we aggregate all eligible news released from 9:00 a.m. on trading day $t-1$ up to 9:00 a.m. on trading day t. News released between 9:00 a.m. and 9:30 a.m. on day t is excluded to allow a minimum 30-minute buffer between the end of the news-collection window and trade execution at the market open (Tetlock, 2007). For news released on weekends or market holidays, we map the next available trading day to day t and the previous trading day to day $t-1$.

We form portfolios at the market open, taken to be 9:30 a.m., on trading day t. The portfolio is then held until the market open on trading day $t+1$, so the realized one-day trading return is measured on an open-to-open basis:

$$R_t^{\text{trade}} = \frac{P_{t+1}^{\text{open}} - P_t^{\text{open}}}{P_t^{\text{open}}}.$$

We consider both equal-weighted and value-weighted implementations for the long and short legs of the strategy. Equal weighting provides a simple and robust assessment of the predictive content of news tone across the firm-size distribution, whereas value weighting places greater emphasis on larger firms and may better reflect implementation considerations such as market depth and trading capacity.

6 Discussion: Synthetic Supervision and the Limitations of LLM-Generated Training Data

Recent research has examined the potential and limitations of using large language models (LLMs) to generate synthetic training data for text classification. A particularly relevant contribution is Li et al. (2023), who identify several structural challenges associated with synthetic data. These include weaker performance relative to real labelled datasets, sensitivity to task subjectivity, limited linguistic diversity, potential bias inherited from the generating model, and distributional differences between synthetic and real text. This section discusses these limitations and explains how the design of the present study addresses them.

6.1 Performance Relative to Real Labelled Data

Li et al. (2023) document that classifiers trained purely on synthetic data often underperform those trained on real labelled datasets in standard NLP benchmarks. Synthetic data therefore frequently serve as a supplement to real training data rather than a full substitute. The present study adopts a different validation criterion. Instead of evaluating models solely through benchmark classification accuracy, we examine whether synthetic supervision alters the economic inference derived from textual signals. Specifically, we test whether sentiment measures constructed from synthetically trained models produce similar relations with stock market reactions to earnings-call disclosures and financial news. The results show that, although classification performance differs across models, the estimated tone–return relations remain economically similar. This approach follows the empirical finance tradition of validating textual measures through market outcomes rather than purely linguistic metrics.

6.2 Task Subjectivity

Li et al. (2023) show that synthetic data perform less reliably in highly subjective classification tasks such as humour detection or sarcasm recognition. Financial sentiment classification differs from such tasks in two important respects. First, the domain is relatively structured: earnings calls and financial news follow professional conventions and contain fewer stylistic ambiguities than informal online discourse. Second, sentiment categories in financial texts (positive, neutral, and negative tone) are defined using widely adopted textual conventions. Accordingly, the analysis focuses on a setting where subjectivity is comparatively limited. Moreover, the scope of the conclusions is restricted to financial sentiment measurement rather than claiming general applicability across all NLP tasks.

6.3 Linguistic Diversity and Stylistic Coverage

Another concern raised by Li et al. (2023) is that synthetic datasets may exhibit limited lexical diversity or stylistic variation, potentially restricting generalisation. Our data-generation protocol mitigates this risk in several ways. Synthetic sentences are generated using multiple frontier LLMs rather than a single model, increasing variation in linguistic style. Generation parameters encourage lexical diversity, and the final dataset is constructed through cross-model consensus filtering. These steps are designed to reduce the influence of the stylistic biases of any individual model. In addition, the empirical analysis evaluates several encoder architectures trained on the same synthetic dataset, allowing us to test whether the findings depend on a particular representation model.

6.4 Model-Specific Bias in Synthetic Labelling

Li et al. (2023) also highlight the possibility that synthetic datasets may encode biases from the generating model. If a single LLM generates and labels the training data, the resulting classifier may reproduce that model's systematic biases. Our design mitigates this risk by separating the generation and tagging stages and by using multi-model consensus tagging. Labels are assigned only when multiple models agree on the classification outcome, thereby reducing the influence of any single model's interpretation of sentiment. In addition, we conduct cross-model comparisons and coefficient-equality tests in the market-reaction regressions. These tests show that sentiment measures derived from different models are highly correlated and produce similar economic estimates, suggesting that synthetic supervision primarily introduces measurement noise rather than systematic bias.

6.5 Distributional Differences Between Synthetic and Real Text

Li et al. (2023) caution that synthetic data may fail to reproduce the full distribution of natural language. Rather than assuming that synthetic sentences perfectly replicate real financial discourse, this study treats synthetic training data as an approximate supervisory signal. The validity of this signal is evaluated through downstream empirical tests using real financial texts, including earnings-call transcripts and Dow Jones Newswires articles. These texts are used exclusively for empirical evaluation rather than for model training. By separating the synthetic training corpus from the real-world application corpus, the analysis provides an external validation of the synthetic supervision approach.

6.6 Remaining Concerns and Robustness Considerations

Despite these safeguards, several concerns identified by Li et al. (2023) may still arise. First, synthetic datasets may omit rare linguistic structures or edge cases present in real financial discourse. Second, the effectiveness of synthetic supervision may depend on the quality of the LLMs used to generate the data. Third, the conclusions may not generalise to tasks requiring deeper contextual reasoning or multi-sentence interpretation. The study addresses these concerns in three ways. First, multiple encoder architectures are evaluated using the same synthetic training corpus, demonstrating that the results are not driven by a particular model specification. Second, the sentiment measures are tested across different textual sources, including earnings calls and financial news, providing evidence that the findings are not tied to a single corpus. Finally, the scope of the contribution is explicitly limited: the results should be interpreted as evidence that synthetic supervision can provide a viable alternative for financial sentiment measurement in asset-pricing applications, rather than as a universal solution for all text classification problems.

6.7 Summary

Overall, the limitations identified by Li et al. (2023) provide an important framework for evaluating synthetic supervision. The design of this study addresses these concerns through multi-model data generation, consensus tagging, cross-architecture robustness tests, and external validation using market-reaction regressions. The findings suggest that while synthetic data may not perfectly replicate real labelled corpora, they can nevertheless produce sentiment measures that yield economically consistent inference in empirical finance settings.

7 Conclusion

This paper studies whether synthetic supervision can serve as a transparent substitute for proprietary labelled data in the fine-tuning of financial language models, without materially altering asset-pricing inference. Framing the question as a measurement-invariance problem, we evaluate synthetic supervision not primarily through benchmark classification performance, but through the stability of the tone–return relation in a standard earnings conference call setting.

The analysis proceeds along three complementary dimensions. First, classification evidence indicates that consensus-labelled synthetic data can produce competitive in-domain performance across multiple encoder architectures, while cross-domain performance on the Financial PhraseBank remains below that of domain-aligned proprietary training. Second, distributional comparisons suggest that transcript-level tone measures constructed from synthetic and proprietary supervision are broadly comparable in dispersion and central tendency. Third, and most importantly, regression estimates of the tone–CAR relation remain positive and statistically significant across model variants, with economically similar magnitudes. Conditional on formal coefficient-equality tests, the evidence is consistent with synthetic supervision behaving as additional measurement noise rather than introducing systematic bias in this setting.

These findings should be interpreted within the deliberately narrow scope of the study. We focus exclusively on sentiment in earnings conference calls and evaluate economic validity within a short-window event-study framework. The results therefore do not imply universal interchangeability between synthetic and proprietary labels across all financial text tasks. Rather, they suggest that, under a transparent and carefully documented consensus protocol, synthetic supervision may preserve economically relevant inference for this particular construct.

More broadly, the study speaks to an emerging structural shift in empirical finance. As large language models become central tools in textual analysis, the provenance of training data increasingly matters for transparency and replicability. Many state-of-the-art financial NLP models rely on proprietary corpora that are inaccessible to researchers, creating barriers to verification and cumulative knowledge building. Synthetic data—when generated, filtered, and documented under explicit protocols—may provide a replicable alternative training substrate. If future research confirms that economically meaningful constructs remain stable under synthetic supervision, this approach could reduce dependence on opaque data pipelines and enhance methodological transparency.

At the same time, the expanding use of generative AI introduces new methodological responsibilities. Synthetic data may embed latent biases inherited from foundation models, amplify stylistic regularities, or underrepresent rare linguistic phenomena. Measurement invariance must therefore be established construct by construct, context by context. The appropriate benchmark is not classification accuracy alone, but the preservation of economically interpretable relationships.

Looking forward, synthetic supervision may become increasingly important in areas where labelled financial data are scarce, sensitive, or costly to obtain, including ESG disclosures, risk-factor narratives, regulatory filings, and cross-lingual financial communication. Advances in prompt design, multi-model consensus, and hybrid human–AI validation may further strengthen the reliability of synthetic labelling pipelines. In parallel, future work could examine dynamic adaptation of synthetic corpora, domain transfer across jurisdictions, and robustness under alternative economic specifications.

In the new era of AI-assisted empirical research, the question is not whether synthetic data perfectly replicate real corpora, but whether they preserve the economic content necessary for valid inference. Conditional on the evidence presented here, synthetic supervision appears capable of sustaining economically stable tone–return relations in a canonical asset-pricing setting. If this pattern generalises, synthetic data may play a significant role in shaping the next generation of transparent, reproducible, and scalable financial textual analysis.

References

Chen, Y., Kelly, B. T., and Xiu, D. (2022). Expected Returns and Large Language Models.

De Kok, T. (2025). ChatGPT for Textual Analysis? How to Use Generative LLMs in Accounting Research. *Management Science*, 71(9):7888–7906.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Frankel, R., Jennings, J., and Lee, J. (2022). Disclosure Sentiment: Machine Learning vs. Dictionary Methods. *Management Science*, 68(7):5514–5532.

Galke, L., Scherp, A., Diera, A., Karl, F., Lin, B. X., Khera, B., Meuser, T., and Singhal, T. (2025). Are We Really Making Much Progress in Text Classification? A Comparative Review.

Huang, A. H., Wang, H., and Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, 40(2):806–841.

Huang, A. H., Zang, A. Y., and Zheng, R. (2014a). Evidence on the Information Content of Text in Analyst Reports. *The Accounting Review*, 89(6):2151–2180.

Huang, X., Teoh, S. H., and Zhang, Y. (2014b). Tone Management. *The Accounting Review*, 89(3):1083–1113.

Li, X., Zhang, Y., Wang, S., and Liu, Y. (2023). Synthetic data generation with large language models for text classification: Potential and limitations. *arXiv preprint arXiv:2305.12345*.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv:1907.11692 [cs].

Loughran, T. and McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2010.01625.x.

Malo, P., Sinha, A., Takala, P., Korhonen, P., and Wallenius, J. (2013). Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts.

Mo, H. and Ouyang, S. (2025). (Generative) AI in Financial Economics. *Journal of Chinese Economic and Business Studies*, 23(4):509–587.

Price, S. M., Doran, J. S., Peterson, D. R., and Bliss, B. A. (2012). Earnings conference calls and stock returns: The incremental informativeness of textual tone. *Journal of Banking & Finance*, 36(4):992–1011.

Rahimikia, E. and Drinkall, F. (2024). Re(Visiting) Large Language Models in Finance.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3):1139–1168. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2007.01232.x.

Table 1: Synthetic Data Construction and Validation of Fine-Tuned Sentiment Models

A. Consensus-Based Construction of the Synthetic Fine-Tuning Corpus

		Counts				Percentages (%)			
Group	Model Label	GPT	Gemini	Mistral	Opus	GPT	Gemini	Mistral	Opus
Matched	Negative	2317	2903	2739	3605	34.35	27.09	35.24	34.38
	Neutral	1038	740	1005	568	15.39	6.90	12.93	5.42
	Positive	2840	5884	3149	5389	42.10	54.90	40.52	51.39
	Total	6195	9527	6893	9562	91.83	88.89	88.69	91.18
Unmatched	Total	551	1191	879	925	8.17	11.11	11.31	8.82

B. Human-Labelled Synthetic Benchmark Corpus and Tagger Agreement Rates

		Counts				Agreement Rate			
Group	Human Label	Cohere	Llama	Palm	Total	Cohere	Llama	Palm	Total
Matched	Pos	5	12	5	22	0.80	0.92	0.80	0.86
	Neu	2	6	7	15	1.00	0.83	1.00	0.93
	Neg	9	4	10	23	1.00	1.00	0.70	0.87
	Total	16	22	22	60	0.94	0.91	0.82	0.88
Unmatched	Pos	2	6	5	13	0.25	0.50	0.50	0.46
	Neu	1	7	3	11	0.50	0.50	0.50	0.50
	Neg	1	8	12	21	0.50	0.50	0.50	0.50
	Total	4	21	20	45	0.38	0.50	0.50	0.49

C. Classification Performance on Internal and External Test Sets

	Overall (%)				Recall (%)		
	Accuracy	Precision	Recall	F_1	Positive	Neutral	Negative
(a) Synthetic benchmark test set							
FinBERT	72.38	65.98	65.32	62.00	82.86	15.38	97.73
FinBERT$_{\text{syn}}$	75.71	77.71	75.84	74.92	67.43	80.77	79.32
BERT$_{\text{syn}}$	76.95	78.59	77.45	76.34	72.00	83.08	77.27
RoBERTa$_{\text{syn}}$	78.86	79.51	78.11	77.31	64.86	80.38	89.09
FinTextBase$_{\text{syn}}$	79.62	80.01	79.18	78.44	84.77	79.62	73.14
FinTextSmall$_{\text{syn}}$	74.00	75.54	74.30	73.50	73.41	76.92	72.57
(b) Financial PhraseBank (External Validation)							
FinBERT	91.70	91.89	88.56	89.71	74.39	98.56	92.74
FinBERT$_{\text{syn}}$	71.20	72.10	71.72	71.19	64.12	72.49	78.55
BERT$_{\text{syn}}$	62.19	64.07	61.41	61.47	53.46	64.97	65.81
RoBERTa$_{\text{syn}}$	76.94	76.34	76.89	76.19	63.72	80.22	86.73
FinTextBase$_{\text{syn}}$	67.77	72.37	77.24	71.18	82.28	56.29	93.14
FinTextSmall$_{\text{syn}}$	66.77	70.80	72.42	68.37	82.05	58.30	76.90

Notes: This table documents the end-to-end synthetic supervision pipeline. Panel A reports the construction of the synthetic fine-tuning corpus using a multi-model consensus protocol; only unanimously labelled sentences are retained for training. Panel B presents the human-labelled synthetic benchmark corpus, which includes both matched (taggers agree) and unmatched (taggers disagree) sentences; the manual label is treated as the ground truth. Agreement rates measure the fraction of the two non-generating tagger labels that coincide with the manual benchmark, aggregated using pooled agreement. Panel C reports classification performance of encoder models fine-tuned on the synthetic corpus, evaluated on (i) the human-labelled synthetic benchmark test set and (ii) the external Financial PhraseBank dataset. Statistics are averaged over 10 random seeds.

Table 2: Descriptive Statistics

Variable	N	Mean	SD	P1	Q1	Median	Q3	P99
CAR_3day	27291	0.040	5.811	-17.444	-3.200	0.085	3.395	16.454
Earn	27291	0.017	0.020	-0.053	0.006	0.014	0.026	0.080
UE	27291	0.073	0.372	-1.861	0.000	0.045	0.148	1.614
Accruals	27291	-0.011	0.024	-0.106	-0.021	-0.010	-0.000	0.062
EarnVol	27291	0.013	0.014	0.000	0.004	0.008	0.016	0.079
Size	27291	9.973	1.360	7.447	8.959	9.830	10.819	14.045
MtoB_assets_q	27291	2.285	1.627	0.886	1.226	1.736	2.672	9.778
Leverage	27291	0.253	0.168	0.000	0.129	0.238	0.350	0.857
IO_PCT	27291	0.782	0.172	0.015	0.717	0.813	0.889	1.058
LogAnalystFollowing	27291	2.831	0.409	1.609	2.565	2.890	3.091	3.611
Age	27291	3.481	0.680	1.674	3.033	3.593	3.967	4.570
CumTurnover	27291	0.596	0.407	0.176	0.340	0.473	0.703	2.533
PriorAlpha	27291	0.007	0.180	-0.506	-0.098	0.011	0.114	0.530
Dividend	27291	0.172	0.377	0.000	0.000	0.000	0.000	1.000
NASDAQ	27291	0.276	0.447	0.000	0.000	0.000	1.000	1.000
Tone_FinBERT	27291	0.252	0.109	-0.012	0.177	0.253	0.329	0.498
Tone_FinBERT$_{syn}$	27291	0.248	0.116	-0.014	0.167	0.246	0.329	0.523
Tone_BERT$_{syn}$	27291	0.287	0.103	0.046	0.216	0.287	0.358	0.530
Tone_RoBERTa$_{syn}$	27291	0.230	0.115	-0.036	0.151	0.230	0.309	0.508
Tone_FinTextBase$_{syn}$	27291	0.288	0.122	0.000	0.203	0.288	0.374	0.571
Tone_FinTextSmall$_{syn}$	27291	0.209	0.114	-0.052	0.129	0.207	0.288	0.483

Notes: This table reports descriptive statistics for the dependent and independent variables used in Table 3, which examines the relationship between short-window market reactions and textual sentiment in earnings conference calls. Variable definitions are provided in Appendix A.

Table 3: Variable definitions

Variable	**Definition**
$Tone_j$	Tone of the earnings conference call measured from sentences spoken by executives during both the presentation and Q&A sections. For model $j \in \{\text{FinBERT}, \text{FinBERT\_syn}, \text{BERT}, \text{RoBERTa}, \text{FinTextBase}, \text{FinTextSmall}\}$, each executive sentence is classified into *positive*, *neutral*, or *negative* using the sentiment class with the highest predicted probability. We then compute $Tone_j = (N_j^+ - N_j^-)/N_j$, where N_j^+ and N_j^- are the numbers of positive and negative executive sentences and N_j is the total number of executive sentences.
CAR	Cumulative abnormal returns in the three-day window around the earnings conference call date, multiplied by 100. The abnormal return equals the raw return minus the value-weighted market return.
$Earn$	Income before extraordinary items in a quarter scaled by the book value of assets at the beginning of the quarter.
UE	Actual quarterly earnings per share (EPS) minus the analyst consensus EPS forecast issued immediately before the earnings announcement date, scaled by the share price at the end of the fiscal quarter and multiplied by 100. Actual EPS and analyst consensus forecasts are from I/B/E/S.
$Accruals$	Net income minus cash flow from operations during the quarter scaled by the book value of assets at the beginning of the quarter.
$EarnVol$	Standard deviation of return on assets during the prior 16 quarters, where return on assets is calculated as quarterly earnings scaled by total assets at the beginning of that quarter.
$Size$	Log of the book value of total assets at the end of the quarter.
$MtoB\_assets\_q$	Market value of equity plus book value of total liabilities scaled by book value of total assets at the end of the quarter.
$Leverage$	Ratio of long-term debt to total assets at the end of the quarter.
IO_PCT	Proportion of outstanding shares held by institutions. Institutional holding data are from Thomson 13F filings database.
$LogAnalystFollowing$	Logarithm of one plus the number of analysts following the firm.
Age	Logarithm of one plus the number of years since a firm first appeared in CRSP monthly file.
$CumTurnover$	Cumulative turnover ratio in the window $[-65, -6]$ relative to the conference call date. Turnover ratio equals the number of shares traded scaled by the number of outstanding shares.
$PriorAlpha$	The intercept from a firm-specific regression of the Fama–French three-factor model using daily data in the window $[-65, -6]$ relative to the conference call date, multiplied by 100.
$Dividend$	An indicator variable that equals one if a firm declares dividends within the three-day window surrounding the conference call date, and zero otherwise.
$NASDAQ$	An indicator variable that equals one if a firm is traded on NASDAQ, and zero otherwise.

Notes: All accounting variables are measured at the quarterly frequency. When applicable, “beginning of the quarter” refers to the start-of-quarter book value of total assets.

Table 4: Quarterly CAR Response to Earnings Call Sentiment

	FinBERT	FinBERT_syn	BERT	RoBERTa	FinTextBase	FinTextSmall
Tone	1.536***	1.476***	1.402***	1.456***	1.496***	1.312***
	(22.58)	(21.11)	(19.18)	(20.84)	(21.40)	(18.37)
Earn	20.816***	20.394***	20.615***	20.227***	19.633***	20.929***
	(5.21)	(5.16)	(5.23)	(5.12)	(4.97)	(5.30)
UE	3.042***	3.140***	3.173***	3.114***	3.125***	3.206***
	(16.19)	(16.46)	(16.47)	(16.21)	(16.44)	(16.52)
Accruals	-19.893***	-19.591***	-19.454***	-19.647***	-19.867***	-19.469***
	(-9.41)	(-9.45)	(-9.25)	(-9.39)	(-9.51)	(-9.32)
EarnVol	-5.255	-3.450	-4.310	-4.460	-3.714	-3.353
	(-1.06)	(-0.70)	(-0.89)	(-0.90)	(-0.76)	(-0.68)
Size	-0.668***	-0.743***	-0.755***	-0.753***	-0.724***	-0.738***
	(-3.81)	(-4.33)	(-4.37)	(-4.41)	(-4.23)	(-4.34)
MtoB	-0.426***	-0.448***	-0.449***	-0.436***	-0.454***	-0.436***
	(-5.94)	(-6.42)	(-6.50)	(-6.29)	(-6.52)	(-6.28)
Leverage	0.916	0.896	0.789	0.817	0.854	0.826
	(1.52)	(1.49)	(1.33)	(1.36)	(1.43)	(1.39)
IO	0.031	0.079	0.077	0.116	0.051	0.113
	(0.05)	(0.13)	(0.13)	(0.20)	(0.09)	(0.20)
AnalystFollowing	-0.479**	-0.405*	-0.495**	-0.446**	-0.399*	-0.466**
	(-2.24)	(-1.90)	(-2.36)	(-2.08)	(-1.89)	(-2.20)
Age	-0.668	-0.616	-0.693	-0.582	-0.605	-0.614
	(-1.36)	(-1.28)	(-1.45)	(-1.20)	(-1.25)	(-1.30)
CumTurnover	1.527***	1.494***	1.445***	1.459***	1.526***	1.459***
	(6.15)	(6.00)	(5.84)	(6.15)	(5.91)	(5.98)
PriorAlpha	-1.546***	-1.450***	-1.362***	-1.434***	-1.476***	-1.305***
	(-6.36)	(-5.95)	(-5.58)	(-5.91)	(-6.06)	(-5.35)
Dividend	0.183	0.182	0.177	0.187	0.194	0.178
	(1.31)	(1.32)	(1.28)	(1.36)	(1.41)	(1.29)
NASDAQ	-1.563***	-1.565***	-1.388***	-1.487***	-1.548***	-1.171***
	(-10.55)	(-10.74)	(-9.43)	(-10.23)	(-10.56)	(-8.41)
Intercept	11.198***	11.449***	11.902***	11.589***	11.236***	11.179***
	(5.00)	(5.16)	(5.40)	(5.26)	(5.09)	(5.11)
Year-Qtr FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,291	27,291	27,291	27,291	27,291	27,291
Adjusted R^2	0.100	0.093	0.090	0.093	0.094	0.086

Notes: The dependent variable is CAR, defined as the three-day cumulative abnormal return around the earnings conference call date. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Tone variables are constructed from executive speech in earnings conference calls and are normalized to have mean zero and unit standard deviation. All regressions include year-quarter and firm fixed effects. All continuous variables are winsorized at the 1st and 99th percentiles. The sample consists of firm-quarter earnings-call observations. *Significance:* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Pairwise Correlations of Executive Tone Measures

	FinBERT	FinBERT_syn	BERT	RoBERTa	FinTextBase	FinTextSmall
FinBERT	1.000	0.935	0.880	0.932	0.929	0.905
FinBERT_syn	0.935	1.000	0.936	0.970	0.980	0.971
BERT	0.880	0.936	1.000	0.935	0.938	0.936
RoBERTa	0.932	0.970	0.935	1.000	0.969	0.958
FinTextBase	0.929	0.980	0.938	0.969	1.000	0.966
FinTextSmall	0.905	0.971	0.936	0.958	0.966	1.000

Notes: This table reports Pearson correlations across transcript-level executive tone measures constructed from executive speech in earnings conference calls. Each tone measure is computed using the corresponding sentiment model indicated by the column/row label. Correlations are computed over the common estimation sample used in the baseline regressions.

Table 6: Pairwise horse-race regressions and Wald tests of coefficient equality

Benchmark	Alternative	$\hat{\beta}_{\text{Bench}}$		$\hat{\beta}_{\text{Alt}}$		p-value: $\beta_B = \beta_A$
		Coef.	t	Coef.	t	
Panel A: Benchmark = FinBERT						
FinBERT	BERT	1.751	(15.27)	0.050	(0.42)	0.000
FinBERT	FinBERT_syn	2.055	(13.46)	-0.317	(-2.11)	0.000
FinBERT	RoBERTa	1.853	(12.44)	-0.077	(-0.53)	0.000
FinBERT	FinTextBase	1.811	(12.39)	-0.027	(-0.18)	0.000
FinBERT	FinTextSmall	2.403	(18.47)	-0.795	(-6.26)	0.000
Panel B: Benchmark = FinBERT_syn						
FinBERT_syn	BERT	1.358	(9.77)	0.406	(2.83)	0.000
FinBERT_syn	RoBERTa	0.712	(3.65)	1.047	(5.43)	0.378
FinBERT_syn	FinTextBase	0.391	(1.73)	1.366	(5.93)	0.030
FinBERT_syn	FinTextSmall	2.849	(14.90)	-1.245	(-6.63)	0.000
Panel C: Benchmark = BERT						
BERT	RoBERTa	0.323	(2.23)	1.434	(10.36)	0.000
BERT	FinTextBase	0.227	(1.51)	1.543	(10.44)	0.000
BERT	FinTextSmall	1.344	(9.65)	0.326	(2.43)	0.000
Panel D: Benchmark = RoBERTa						
RoBERTa	FinTextBase	0.697	(3.56)	1.076	(5.31)	0.332
RoBERTa	FinTextSmall	2.560	(15.15)	-0.952	(-5.64)	0.000
Panel E: Benchmark = FinTextBase						
FinTextBase	FinTextSmall	2.934	(14.86)	-1.327	(-6.94)	0.000

Year-quarter and firm fixed effects included; standard errors clustered by firm.
N = 27,291 in all regressions; adjusted R^2 ranges from 0.050 to 0.064.

Notes: Each row reports a two-tone "horse-race" regression in which both the benchmark and alternative tone measures enter jointly, along with the full set of controls. Tone variables are standardized within the estimation sample. The final column reports the p-value from a Wald test of coefficient equality $H_0 : \beta_{\text{Bench}} = \beta_{\text{Alt}}$.

Table 7: Subsample stability: interaction regressions

	FinBERT	FinBERT_syn	BERT	RoBERTa	FinTextBase	FinTextSmall
Panel A: Firm size (Big = 1)						
β (Small firms)	1.782	1.704	1.678	1.703	1.731	1.533
	(20.82)	(18.82)	(17.82)	(19.12)	(19.16)	(16.57)
Δ (Big−Small)	−0.519	−0.470	−0.551	−0.508	−0.486	−0.450
	(-4.77)	(-4.21)	(-4.93)	(-4.56)	(-4.39)	(-3.99)
p-value: $\Delta = 0$	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: Analyst coverage (HighCov = 1)						
β (Low coverage)	1.612	1.551	1.485	1.543	1.566	1.398
	(19.75)	(18.42)	(17.26)	(18.22)	(18.65)	(16.31)
Δ (High−Low)	−0.143	−0.145	−0.157	−0.167	−0.134	−0.167
	(-1.43)	(-1.45)	(-1.55)	(-1.67)	(-1.34)	(-1.71)
p-value: $\Delta = 0$	0.153	0.148	0.211	0.096	0.179	0.093
Panel C: Time period (Post2020 = 1)						
β (Pre-2020)	1.569	1.505	1.452	1.496	1.529	1.359
	(21.32)	(20.02)	(18.69)	(19.82)	(20.33)	(17.60)
Δ (Post−Pre)	−0.116	−0.103	−0.173	−0.136	−0.113	−0.156
	(-1.25)	(-1.13)	(-1.76)	(-1.50)	(-1.21)	(-1.71)
p-value: $\Delta = 0$	0.210	0.260	0.077	0.134	0.225	0.087
Observations	27,291	27,291	27,291	27,291	27,291	27,291
Firm FE / Year-Qtr FE	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes

Notes: Each column reports a pooled interaction regression of $CAR[-1,+1]$ on a standardized tone measure, controls, firm fixed effects, and year-quarter fixed effects. "Big" ("HighCov") equals one for observations above the year-quarter median of log book value of total assets (log analyst following). "Post2020" equals one for conference calls on or after 2020-01-01. β is the tone loading in the baseline group; Δ is the incremental loading for the indicated group. p-values are from Wald tests of $H_0 : \Delta = 0$. Standard errors are clustered by firm.

Table 8: Sample Construction of Dow Jones News Data

Sample stage	Number of observations
NewsRaw	20,761,104
NewsChained	14,599,698
NewsFiltered	3,693,711
Newswith1Stock	1,399,262
NewsDeduplicate	1,182,441
Headline only	20,430
Headline + body story	1,162,011

Notes: This table reports the number of observations remaining after each stage of the news-sample construction procedure. `NewsRaw` is the full Dow Jones Newswires Text News Feed from January 2013 to December 2023. `NewsChained` denotes the sample after consolidating raw feed records into economically meaningful news items. `NewsFiltered` retains only news items related to exactly one stock. `Newswith1Stock` further requires a valid one-day open-to-open stock return and market capitalization. `NewsDeduplicate` excludes articles whose bag-of-words cosine similarity is at least 0.8 relative to an earlier article published within the prior five business days. The final two rows report the composition of `NewsDeduplicate` by text availability.

Table 9: Performance of Equal-Weighted News-Tone Portfolios

Quintile	FinBERT	FinBERT_syn	BERT	RoBERTa	FinTextSmall	FinTextBase
Panel A. Quintile portfolio returns						
1	-2.52%	-7.56%	-10.08%	-15.12% *	-7.56%	-12.60%
	(-0.50)	(-0.95)	(-1.45)	(-1.93)	(-0.93)	(-1.60)
2	17.64% **	17.64% **	15.12% **	20.16% ***	15.12% **	15.12% **
	(2.40)	(2.31)	(2.22)	(2.70)	(2.12)	(2.25)
3	25.20% ***	30.24% ***	37.80% ***	32.76% ***	35.28% ***	35.28% ***
	(3.57)	(4.11)	(5.13)	(4.46)	(4.64)	(4.84)
4	40.32% ***	42.84% ***	40.32% ***	47.88% ***	42.84% ***	42.84% ***
	(5.66)	(5.92)	(5.63)	(6.20)	(5.80)	(6.00)
5	37.80% ***	35.28% ***	32.76% ***	32.76% ***	32.76% ***	32.76% ***
	(5.55)	(5.22)	(4.91)	(4.89)	(4.81)	(5.03)
Panel B. Long–short portfolio performance						
5–1 Return	40.61% ***	41.56% ***	43.16% ***	46.75% ***	38.73% ***	45.39% ***
	(8.53)	(8.98)	(9.33)	(9.94)	(8.50)	(9.67)
Sharpe Ratio	2.5995	2.7719	2.8864	3.1592	2.5615	3.0119
Excess returns	39.35% ***	40.54% ***	42.34% ***	45.89% ***	38.18% ***	44.20% ***
	(8.32)	(8.80)	(9.12)	(9.72)	(8.39)	(9.48)
One-factor alpha	39.92% ***	40.58% ***	42.34% ***	46.02% ***	38.13% ***	44.15% ***
	(8.42)	(8.75)	(9.07)	(9.81)	(8.38)	(9.44)
Three-factor alpha	39.42% ***	40.29% ***	42.06% ***	45.65% ***	38.07% ***	43.95% ***
	(8.49)	(8.78)	(9.13)	(9.88)	(8.40)	(9.50)
Four-factor alpha	39.24% ***	40.17% ***	41.94% ***	45.50% ***	37.94% ***	43.77% ***
	(8.49)	(8.75)	(9.09)	(9.83)	(8.37)	(9.44)
Five-factor alpha	38.96% ***	40.04% ***	41.61% ***	45.24% ***	37.63% ***	43.39% ***
	(8.36)	(8.68)	(9.02)	(9.84)	(8.29)	(9.37)
Panel C. Long–short portfolio performance with transaction costs						
5–1 Return	-5.50%	-4.58%	-2.99%	0.85%	-6.71%	-0.60%
	(-1.15)	(-0.99)	(-0.65)	(0.18)	(-1.47)	(-0.13)
Sharpe Ratio	-0.3519	-0.3051	-0.1999	0.0573	-0.4441	-0.0401
Excess returns	-6.73%	-5.56%	-3.79%	-0.00%	-7.40%	-1.79%
	(-1.42)	(-1.21)	(-0.82)	(-0.00)	(-1.63)	(-0.38)
One-factor alpha	-6.17%	-5.53%	-3.78%	0.12%	-7.45%	-1.84%
	(-1.30)	(-1.19)	(-0.81)	(0.03)	(-1.64)	(-0.39)
Three-factor alpha	-6.67%	-5.81%	-4.07%	-0.25%	-7.51%	-2.04%
	(-1.43)	(-1.27)	(-0.88)	(-0.05)	(-1.66)	(-0.44)
Four-factor alpha	-6.85%	-5.94%	-4.18%	-0.40%	-7.64%	-2.22%
	(-1.48)	(-1.29)	(-0.91)	(-0.09)	(-1.69)	(-0.48)
Five-factor alpha	-7.13%	-6.07%	-4.51%	-0.65%	-7.94%	-2.61%
	(-1.53)	(-1.32)	(-0.98)	(-0.14)	(-1.75)	(-0.56)

This table reports the performance of equal-weighted portfolios formed by sorting stocks each trading day into quintiles based on the firm-level news-tone signal. Panel A reports the annualized raw returns of quintile portfolios, where Quintile 1 denotes the lowest-tone portfolio and Quintile 5 denotes the highest-tone portfolio. Panel B reports the annualized return and Sharpe ratio of the long–short portfolio formed by buying Quintile 5 and shorting Quintile 1, together with the annualized excess return and annualized factor-adjusted alphas from the one-factor, three-factor, four-factor, and five-factor models. Panel C reports the corresponding long–short results after accounting for turnover-based transaction costs, where the daily transaction cost equals 10 basis points times the long–short portfolio turnover. Returns and alphas are annualized by multiplying daily estimates by 252. All return and alpha entries are reported in percentage points and marked with %. Returns and alphas are annualized by multiplying daily estimates by 252. Numbers in parentheses are Newey–West t-statistics with 4 lags. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 10: Synthetic Fine-Tuning Datasets for Other Categories

	Counts				Percentages (%)			
	GPT	Gemini	Mistral	Opus	GPT	Gemini	Mistral	Opus
(a) ESG4								
Unanimous	2300	2981	2743	2752	95.87	88.77	94.68	84.21
2-vs-1	98	353	133	425	4.09	10.51	4.59	13.00
All-different	1	24	21	91	0.04	0.71	0.72	2.78
Environmental	611	547	637	572	25.47	16.29	21.99	17.50
Social	519	663	691	646	21.63	19.74	23.85	19.77
Governance	546	512	545	602	22.76	15.25	18.81	18.42
Non-ESG	624	1259	870	932	26.01	37.49	30.03	28.52
(b) ESG11								
Unanimous	8085	10050	10765	8296	94.89	90.24	95.94	87.59
2-vs-1	418	1010	436	1061	4.91	9.07	3.89	11.20
All-different	17	77	19	114	0.20	0.69	0.17	1.20
Climate Change	695	679	925	611	8.16	6.10	8.24	6.45
Corp Behavior	537	554	551	523	6.30	4.97	4.91	5.52
Corp Governance	747	1022	973	799	8.77	9.18	8.67	8.44
Env Opportunities	833	909	879	818	9.78	8.16	7.83	8.64
Human Capital	855	1090	1015	1026	10.04	9.79	9.05	10.83
Natural Capital	807	768	807	608	9.47	6.90	7.19	6.42
Non-ESG	877	2058	2479	1332	10.29	18.48	22.09	14.06
Pollution and Waste	805	764	806	611	9.45	6.86	7.18	6.45
Product Liability	775	965	1088	797	9.10	8.66	9.70	8.42
Social Opportunities	637	729	738	595	7.48	6.55	6.58	6.28
Stakeholder Oppos	517	512	504	576	6.07	4.60	4.49	6.08
(c) Emotions								
Unanimous	6189	6823	5775	5186	92.19	85.55	96.46	88.42
2-vs-1	492	1097	198	630	7.33	13.76	3.31	10.74
All-different	32	55	14	49	0.48	0.69	0.23	0.84
Anger	514	557	625	509	7.66	6.98	10.44	8.68
Anticipation	858	1037	903	742	12.78	13.00	15.08	12.65
Disgust	512	719	676	528	7.63	9.02	11.29	9.00
Fear	859	1074	860	739	12.80	13.47	14.36	12.60
Joy	853	1016	780	728	12.71	12.74	13.03	12.41
Sadness	943	811	526	663	14.05	10.17	8.79	11.30
Surprise	787	515	686	504	11.72	6.46	11.46	8.59
Trust	863	1094	719	773	12.86	13.72	12.01	13.18
(d) Forward-looking Statement (FLS)								
Unanimous	1758	1811	1638	1895	93.11	85.55	93.17	67.82
2-vs-1	128	296	117	829	6.78	13.98	6.66	29.67
All-different	2	10	3	70	0.11	0.47	0.17	2.51
Specific FLS	648	609	543	785	34.32	28.77	30.89	28.10
Non-specific FLS	606	691	589	610	32.10	32.64	33.50	21.83
Not FLS	504	511	506	500	26.69	24.14	28.78	17.90
(e) Overconfidence								
Unanimous	1060	1780	1702	1475	97.52	88.29	88.10	88.75
2-vs-1	27	236	230	187	2.48	11.71	11.90	11.25
Overconfidence	514	540	502	500	47.29	26.79	25.98	30.08
Non-overconfidence	546	1240	1200	975	50.23	61.51	62.11	58.66

Notes: Each row reports the count and the percentage of sentences used for fine-tuning under each of the four models. "Unanimous" means all tagger models assign the same label; "2-vs-1" means two taggers agree and one disagrees; "All-different" means all three taggers disagree. The values for each category are reported based on "Unanimous" classification.

Table 11: Human-Labelled Synthetic Benchmark Corpus for Other categories

Task	Manual label	Matched								Unmatched							
		Counts				Agreement Rate				Counts				Agreement Rate			
		Cohere	Llama	Palm	Total	Cohere	Llama	Palm	Total	Cohere	Llama	Palm	Total	Cohere	Llama	Palm	Total
ESG4	Environmental	5	12	4	21	1.000	0.833	1.000	0.905	0	1	2	3	–	0.500	0.250	0.333
	Social	6	8	7	21	0.667	1.000	0.857	0.857	5	7	1	13	0.444	0.308	0.500	0.375
	Governance	7	10	3	20	0.857	0.900	1.000	0.900	4	8	4	16	0.500	0.400	0.250	0.387
	Non-ESG	6	5	7	18	1.000	1.000	1.000	1.000	7	17	24	48	0.667	0.444	0.489	0.494
	Total	24	35	21	80	0.875	0.914	0.952	0.912	16	33	31	80	0.538	0.404	0.443	0.444
ESG11	Climate Change	9	4	7	20	1.000	1.000	1.000	1.000	15	11	28	54	0.400	0.450	0.429	0.425
	Corporate Behavior	11	5	4	20	0.909	1.000	1.000	0.950	23	9	32	64	0.286	0.176	0.297	0.276
	Corporate Governance	7	9	6	22	0.857	0.889	0.833	0.864	25	10	15	50	0.273	0.389	0.345	0.325
	Environmental Opportunities	7	4	11	22	0.857	1.000	0.818	0.864	8	9	19	36	0.333	0.647	0.447	0.471
	Human Capital	7	8	6	21	0.857	1.000	0.833	0.905	9	16	19	44	0.562	0.516	0.405	0.476
	Natural Capital	8	11	1	20	1.000	1.000	1.000	1.000	7	21	18	46	0.417	0.643	0.417	0.522
	Non-ESG	5	8	4	17	1.000	0.875	1.000	0.941	18	16	22	56	0.810	0.619	0.425	0.573
	Pollution and Waste	8	9	2	19	1.000	1.000	1.000	1.000	15	12	24	51	0.433	0.609	0.458	0.485
	Product Liability	8	4	9	21	0.750	1.000	1.000	0.905	23	25	47	95	0.333	0.449	0.366	0.382
	Social Opportunities	9	4	5	18	0.889	1.000	0.800	0.889	8	5	18	31	0.429	0.400	0.361	0.383
	Stakeholder Opposition	3	5	12	20	0.667	1.000	0.917	0.900	11	14	53	78	0.318	0.500	0.377	0.390
	Total	82	71	67	220	0.902	0.972	0.910	0.927	162	148	295	605	0.395	0.507	0.388	0.418
Emo	Anger	6	6	8	20	0.833	1.000	0.625	0.800	4	7	12	23	0.375	0.286	0.208	0.261
	Anticipation	1	8	8	17	0.000	0.750	1.000	0.824	4	57	16	77	0.250	0.278	0.500	0.318
	Disgust	9	1	7	17	0.778	1.000	1.000	0.882	6	14	5	25	0.500	0.269	0.625	0.391
	Fear	7	10	6	23	0.857	0.600	1.000	0.783	6	37	15	58	0.333	0.281	0.407	0.320
	Joy	11	6	8	25	0.636	1.000	0.750	0.760	12	15	9	36	0.458	0.481	0.389	0.449
	Sadness	8	3	7	18	1.000	0.667	1.000	0.944	3	22	14	39	0.500	0.372	0.357	0.377
	Surprise	9	2	10	21	1.000	1.000	0.900	0.952	13	9	11	33	0.385	0.389	0.455	0.409
	Trust	11	2	6	19	1.000	1.000	0.833	0.947	4	20	4	28	0.167	0.444	0.250	0.380
	Total	62	38	60	160	0.855	0.816	0.883	0.856	52	181	86	319	0.392	0.332	0.390	0.358
FLS	Specific FLS	5	12	5	22	0.800	0.833	1.000	0.864	6	8	7	21	0.250	0.438	0.500	0.405
	Non-specific FLS	5	13	2	20	1.000	0.923	1.000	0.950	1	5	2	8	0.500	0.556	0.500	0.533
	Not FLS	6	6	6	18	0.833	1.000	1.000	0.944	9	5	2	16	0.562	0.300	0.500	0.467
	Total	16	31	13	60	0.875	0.903	1.000	0.917	16	18	11	45	0.433	0.429	0.500	0.448
Overc	Overconfidence	3	8	5	16	1.000	1.000	1.000	1.000	2	2	0	4	0.667	0.500	–	0.571
	Non-overconfidence	10	10	4	24	0.900	0.700	1.000	0.833	9	6	1	16	0.529	0.500	0.500	0.516
	Total	13	18	9	40	0.923	0.833	1.000	0.900	11	8	1	20	0.550	0.500	0.500	0.526

Notes: Each sentence is generated by exactly one model (Cohere, Llama, or Palm) and tagged by the other two models. Columns report results separately for cases in which the two tagging models agree on the label (*Matched*) versus disagree (*Unmatched*). For both matched and unmatched samples, the manual label is treated as the true and final label. *Agreement Rate* uses the manual label as the benchmark and is computed as pooled agreement: within each task–label cell, we sum the number of tagger labels equal to the manual label and divide by the total number of available tagger labels (typically two per sentence). "Total" rows aggregate across manual-label categories within each task. Cells with zero sentences imply undefined agreement rates and are reported as "–".

Table 12: Classification Performance of Fine-Tuned Models in Other Categories

	Accuracy	Precision	Recall	F_1
		(a) ESG4		
FinBERT	74.38	71.29	72.13	71.23
$\text{FinBERT}_{\text{syn}}$	77.13	76.07	73.28	73.48
BERT_{syn}	78.31	76.55	74.62	74.51
$\text{RoBERTa}_{\text{syn}}$	78.56	77.19	74.91	74.95
$\text{FinTextBase}_{\text{syn}}$	77.94	77.30	74.11	74.40
$\text{FinTextSmall}_{\text{syn}}$	77.58	77.03	73.70	74.02
		(b) ESG11		
$\text{FinBERT}_{\text{syn}}$	73.16	75.77	73.07	72.37
BERT_{syn}	73.81	75.90	73.71	72.94
$\text{RoBERTa}_{\text{syn}}$	73.81	75.88	73.71	72.75
$\text{FinTextBase}_{\text{syn}}$	73.40	76.04	73.19	72.46
$\text{FinTextSmall}_{\text{syn}}$	73.13	75.32	72.82	72.14
		(c) Emotions		
$\text{FinBERT}_{\text{syn}}$	60.50	62.00	62.24	60.84
BERT_{syn}	65.23	66.83	66.21	65.52
$\text{RoBERTa}_{\text{syn}}$	66.77	68.18	67.62	66.89
$\text{FinTextBase}_{\text{syn}}$	63.43	64.42	64.76	63.51
$\text{FinTextSmall}_{\text{syn}}$	60.82	61.81	62.13	61.04
		(d) Forward-looking statement		
FinBERT	51.43	69.47	52.57	48.19
$\text{FinBERT}_{\text{syn}}$	76.95	78.04	77.73	77.61
BERT_{syn}	74.48	76.45	76.06	75.33
$\text{RoBERTa}_{\text{syn}}$	76.00	78.07	77.26	76.98
$\text{FinTextBase}_{\text{syn}}$	76.11	77.47	77.08	76.88
$\text{FinTextSmall}_{\text{syn}}$	76.10	77.20	76.78	76.78
		(e) Overconfidence		
$\text{FinBERT}_{\text{syn}}$	62.67	66.84	68.00	62.40
BERT_{syn}	59.17	67.16	66.50	58.66
$\text{RoBERTa}_{\text{syn}}$	66.00	69.83	71.37	65.72
$\text{FinTextBase}_{\text{syn}}$	65.09	70.04	71.09	64.83
$\text{FinTextSmall}_{\text{syn}}$	60.78	67.45	67.62	60.59

Notes: This table shows the classification performance (overall Accuracy, Precision, Recall, and F_1) of the fine-tuned FinBERT in Huang et al. (2023) and FinBERT, BERT, RoBERTa, FinText-Base-2023 and FinText-Small-2023 in Rahimikia and Drinkall (2024) ($\text{FinTextBase}_{\text{syn}}$ and $\text{FinTextSmall}_{\text{syn}}$) fine-tuned by our synthetic data. Panel (a) to (e) report, respectively, the performance of each Fine-tuned LLMs on simulated ESG4, ESG11, Emotions, Forward looking statement and Overconfidence simulated data.

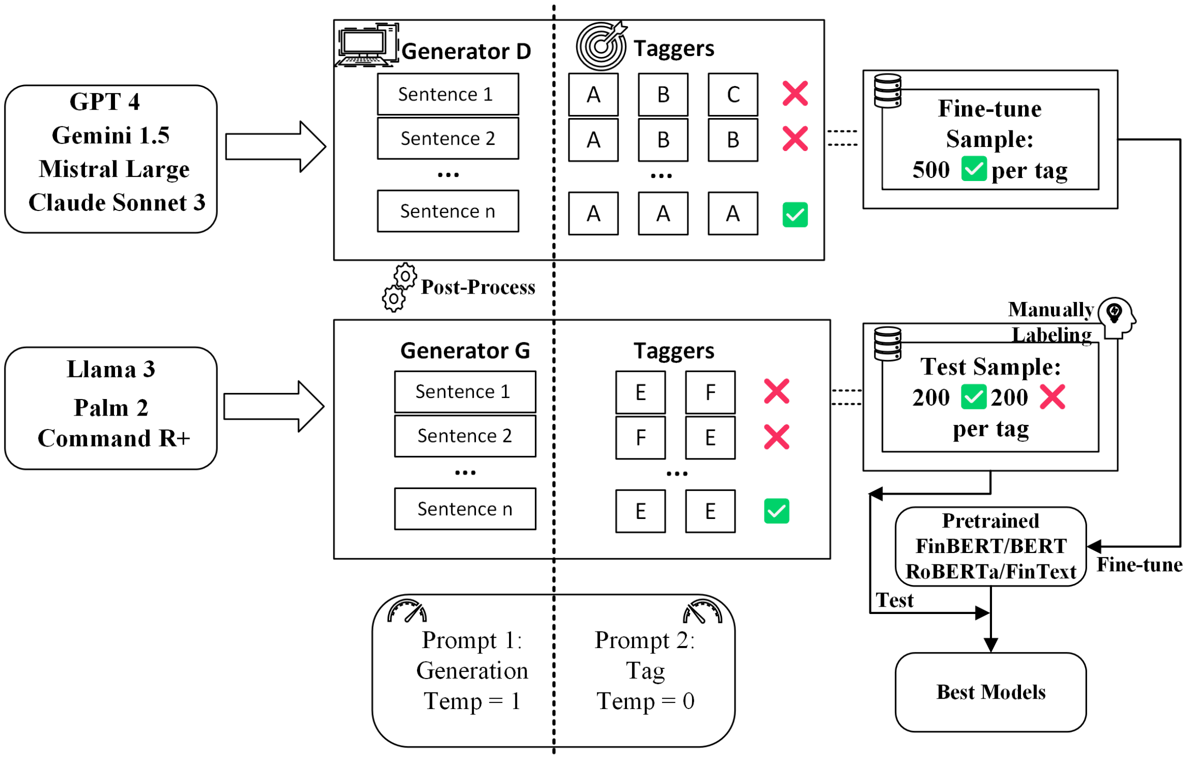


Figure 1: Synthetic Data Generation and Fine-Tuning Workflow

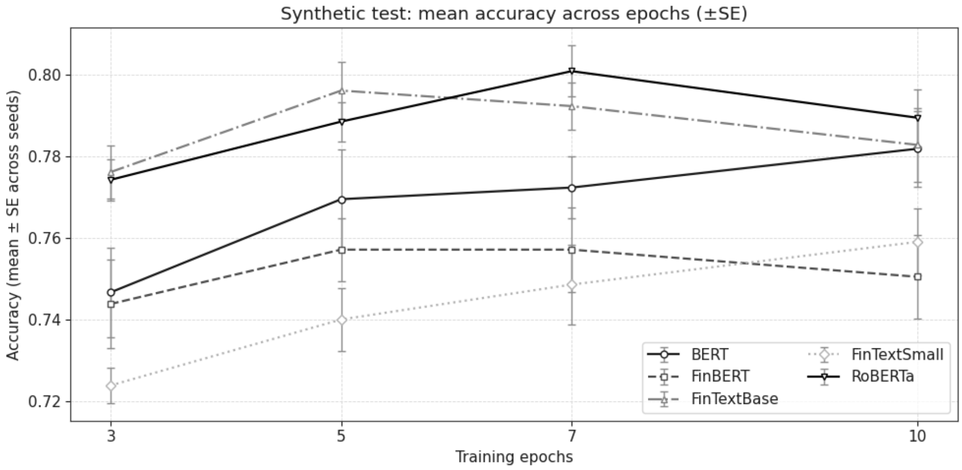


Figure 2: Epoch search on the synthetic benchmark sample. The figure reports mean classification accuracy across 10 random seeds for each encoder backbone at training epochs $\{3, 5, 7, 10\}$, with error bars denoting $\pm$ one standard error across seeds. FinText-Base and FinText-Small correspond to the 2023 model variants (trained on data through 2023).

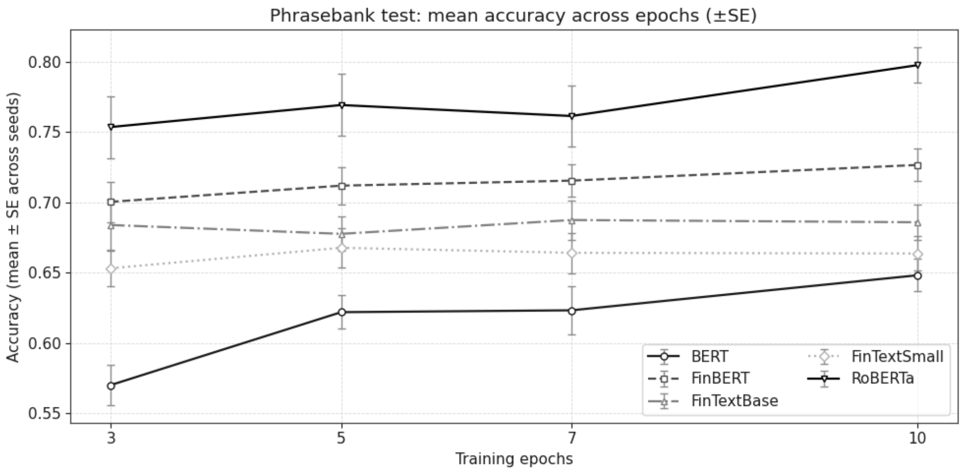
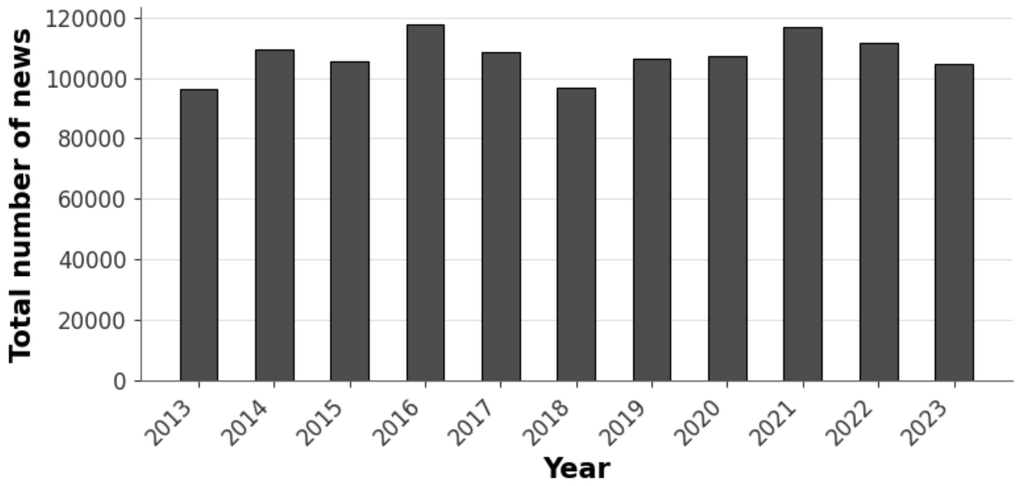
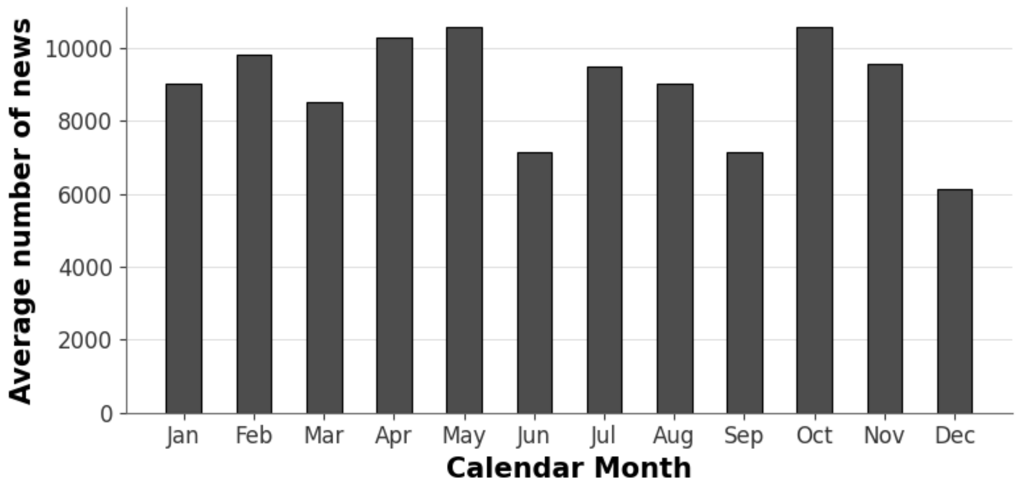


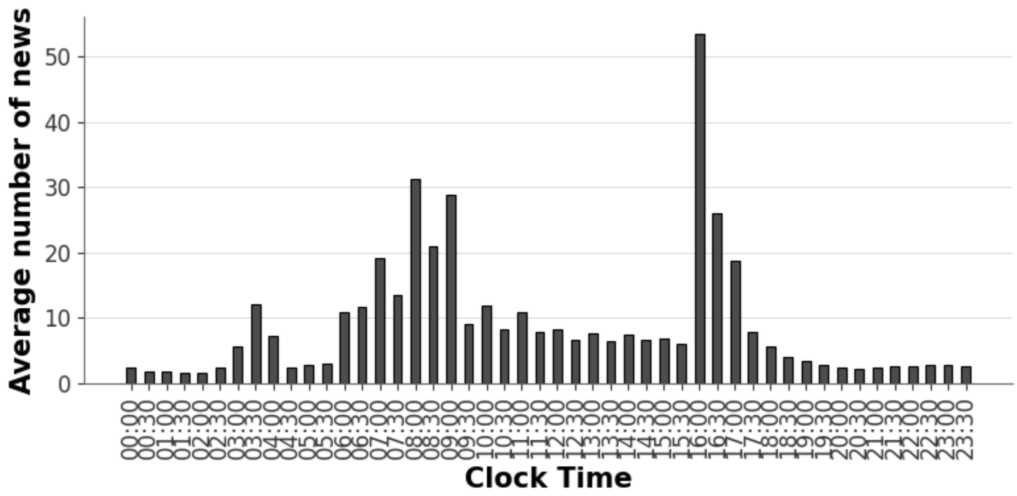
Figure 3: Epoch search on the Financial PhraseBank test set. The figure reports mean accuracy across 10 random seeds for each encoder backbone at training epochs $\{3, 5, 7, 10\}$, with error bars denoting $\pm$ one standard error across seeds. FinText-Base and FinText-Small correspond to the 2023 model variants (trained on data through 2023).



(a) Total number of news by year



(b) Average number of news by calendar month



(c) Average number of news by clock time

Figure 4: Descriptive distribution of Dow Jones news over time. Panel (a) reports the total number of news items by year. Panel (b) reports the average number of news items by calendar month. Panel (c) reports the average number of news items by half-hour clock-time interval.

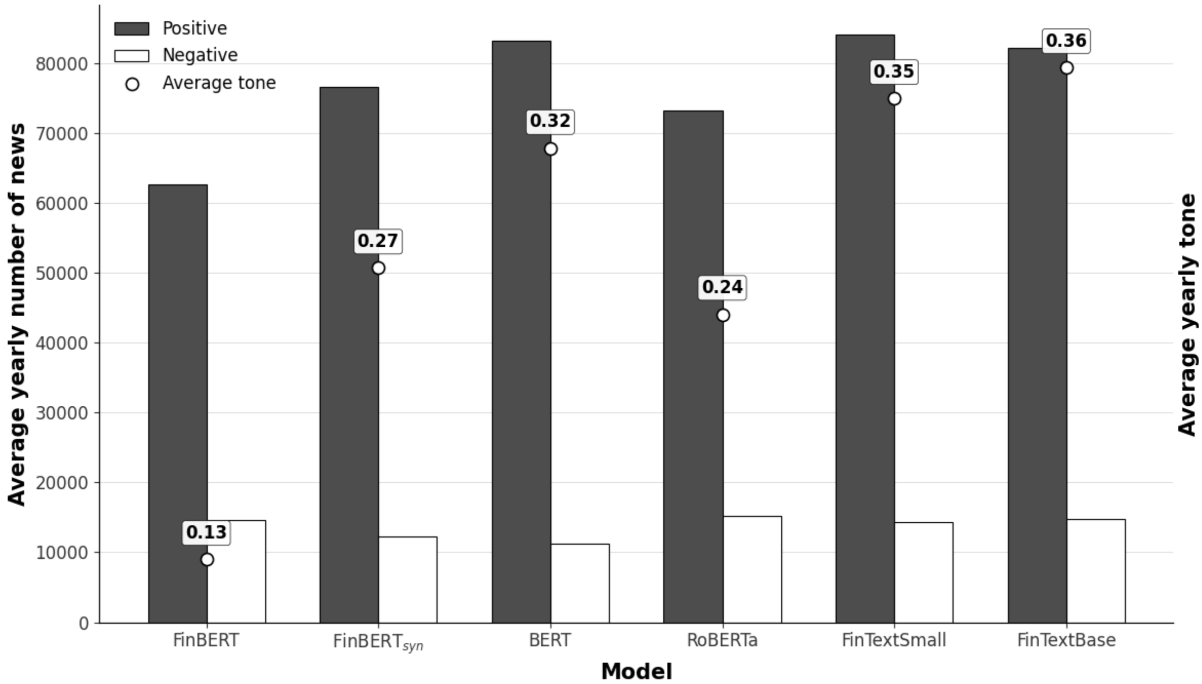


Figure 5: Average yearly number of positive and negative news articles and average yearly tone across sentiment classifiers

On-line Appendix:

A Prompt Engineering Details and Full Prompts

This appendix documents the exact prompt templates used in our two-stage framework. We use (i) a **Stage 1 generation prompt** with temperature = 1 to elicit diverse finance-domain sentences, and (ii) a **Stage 2 tagging prompt** with temperature = 0 to produce deterministic labels. The core structure is held fixed across tasks, and task-specific label definitions are substituted modularly. :contentReference[oaicite:10]index=10

A.1 Sentiment (Positive / Negative / Neutral)

Stage 1: Generation (Temp = 1)

```
You are an expert with extensive experience in various fields of
finance, accounting, banking, business, management, marketing, and
economics. Create a list of 15 sentences on these fields. The sentences
should be structured according to the conventional formats used in news
articles, financial reports, audit reports, financial analyst reports, ESG
and CSR reports, market research reports, business plans, economic analysis
reports, transcripts of conference calls, books, and other documents and
reports relevant to these fields. Explore various sentence structures in
your response. For example, transitional phrases may be utilized at the
beginning of sentences to reflect the continuation and flow typical in
mid-discussion segments of these documents. Create an equal number of
sentences depicting positive, negative, and neutral sentiments, with
```

5 examples of each sentiment. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (positive, negative, or neutral).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them according to the sentiments as either 'Positive', 'Negative', or 'Neutral' sentiments. The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as 'None'. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

A.2 Emotion (Anger / Anticipation / Disgust / Fear / Joy / Sadness / Surprise / Trust)

Stage 1: Generation (Temp = 1)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. Create a list of 40 sentences on these fields. The sentences should be structured according to the conventional formats used in news articles, financial reports, audit reports, financial analyst reports, ESG and CSR reports, market research reports, business plans, economic analysis reports, transcripts of conference calls, books, and other documents and reports relevant to these fields. Explore various sentence structures in your response. For example, transitional phrases may be utilized at the beginning of sentences to reflect the continuation and flow typical in mid-discussion segments of these documents. Create an equal number of sentences depicting anger, anticipation, disgust, fear, joy, sadness, surprise, and trust emotions, with 5 examples of each emotion. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (anger, anticipation, disgust, fear, joy, sadness, surprise, or trust).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them according to the emotions as either ‘Anger’, ‘Anticipation, ‘Disgust’, ‘Fear, ‘Joy’, ‘Sadness’, ‘Surprise’, or ‘Trust’ emotions. The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as ‘None’. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

A.3 ESG(4): Environmental / Social / Governance / Non-ESG

Stage 1: Generation (Temp = 1)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. Create a list of 20 sentences on these fields. The sentences should be structured according to the conventional formats used in news articles, financial reports, audit reports, financial analyst reports, ESG and CSR reports, market research reports, business plans, economic analysis reports, transcripts of conference calls, books, and other documents and reports relevant to these fields. Explore various sentence structures in your response. For example, transitional phrases may be utilized at the beginning of sentences to reflect the continuation and flow typical in mid-discussion segments of these documents. Create an equal number of sentences depicting environmental (which includes: climate change, natural capital, pollution and waste, and environmental opportunities), social (which includes: human capital, product liability, stakeholder opposition, and social opportunities), governance (which includes: corporate governance, and corporate behavior), and non-ESG factors, with 5 examples of each factor. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (environmental, social, governance, or non-ESG).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them according to the ESG factors as either ‘Environmental’ (which includes: climate change, natural capital, pollution and waste, and environmental opportunities), ‘Social’ (which includes: human capital, product liability, stakeholder opposition, and social opportunities), ‘Governance’ (which includes: corporate governance, and corporate behavior), or

‘Non-ESG’ factors. The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as ‘None’. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

A.4 ESG(11): Climate Change / Natural Capital / Pollution and Waste / Environmental Opportunities / Human Capital / Product Liability / Stakeholder Opposition / Social Opportunities / Corporate Governance / Corporate Behavior / Non-ESG

Stage 1: Generation (Temp = 1)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. Create a list of 55 sentences on these fields. The sentences should be structured according to the conventional formats used in news articles, financial reports, audit reports, financial analyst reports, ESG and CSR reports, market research reports, business plans, economic analysis reports, transcripts of conference calls, books, and other documents and reports relevant to these fields. Explore various sentence structures in your response. For example, transitional phrases may be utilized at the beginning of sentences to reflect the continuation and flow typical in mid-discussion segments of these documents. Create an equal number of sentences depicting climate change (which includes: carbon emissions, climate change vulnerability, financing environmental impact, and product carbon footprint), natural capital (which includes: biodiversity and land use, raw material sourcing, and water stress), pollution and waste (which includes: electronic waste, packaging material and waste, and toxic emissions and waste), environmental opportunities (which includes: opportunities in clean tech, opportunities in green building, and opportunities in renewable energy), human capital (which includes: health and safety, human capital development, labor management, and supply chain labor standards), product liability (which includes: chemical safety, consumer financial protection, privacy and data security, product safety and quality, and responsible investment), stakeholder opposition (which includes: community relations, and controversial sourcing), social opportunities (which includes: access to finance, access to health care, and opportunities in nutrition and health), corporate governance (which includes: board, pay, ownership and control, and accounting), corporate behavior (which includes: business ethics, and tax transparency), and non-ESG factors, with 5 examples of each factor. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (climate change, natural capital, pollution and waste, environmental opportunities, human capital, product liability, stakeholder opposition, social opportunities, corporate governance, corporate behavior, or non-ESG).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them according to the ESG factors as either 'Climate Change' (which includes: carbon emissions, climate change vulnerability, financing environmental impact, and product carbon footprint), 'Natural Capital' (which includes: biodiversity and land use, raw material sourcing, and water stress), 'Pollution and Waste' (which includes: electronic waste, packaging material and waste, and toxic emissions and waste), 'Environmental Opportunities' (which includes: opportunities in clean tech, opportunities in green building, and opportunities in renewable energy), 'Human Capital' (which includes: health and safety, human capital development, labor management, and supply chain labor standards), 'Product Liability' (which includes: chemical safety, consumer financial protection, privacy and data security, product safety and quality, and responsible investment), 'Stakeholder Opposition' (which includes: community relations, and controversial sourcing), 'Social Opportunities' (which includes: access to finance, access to health care, and opportunities in nutrition and health), 'Corporate Governance' (which includes: board, pay, ownership and control, and accounting), 'Corporate Behavior' (which includes: business ethics, and tax transparency), or 'Non-ESG' factors. The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as 'None'. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

A.5 Forward-Looking Statements (Specific FLS / Non-specific FLS / Not FLS)

Stage 1: Generation (Temp = 1)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. Create a list of 15 sentences on these fields. The sentences should be structured according to the conventional formats used in news articles, financial reports, audit reports, financial analyst reports, ESG and CSR reports, market research reports, business plans, economic analysis reports, transcripts of conference calls, books, and other documents and reports relevant to these fields. Explore various sentence structures in your response. For example, transitional phrases may be utilized at the beginning of sentences to reflect the continuation and flow typical in mid-discussion segments of these documents.

Create an equal number of sentences depicting ‘Specific FLS’ (which are sentences that concern the future of the company, like plans or prospects), ‘Non-specific FLS’ (which are sentences that concern the future in general or are common to any company, like cautionary remarks, risk disclosures, or market trends), and ‘Not FLS’ (which are sentences that do not concern the future of the company), with 5 examples of each. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (Specific FLS, Non-specific FLS, or Not FLS).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them as either ‘Specific FLS’ (which are sentences that concern the future of the company, like plans or prospects), ‘Non-specific FLS’ (which are sentences that concern the future in general or are common to any company, like cautionary remarks, risk disclosures, or market trends) or ‘Not FLS’ (which are sentences that do not concern the future of the company). The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as ‘None’. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

A.6 Overconfidence (Overconfidence / Non-overconfidence)

Stage 1: Generation (Temp = 1)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. Create a list of 20 sentences on these fields. The sentences should be structured according to the conventional formats used in news articles, financial reports, audit reports, financial analyst reports, ESG and CSR reports, market research reports, business plans, economic analysis reports, transcripts of conference calls, books, and other documents and reports relevant to these fields. Explore various sentence structures in your response. For example, transitional phrases may be utilized at the beginning of sentences to reflect the continuation and flow typical in mid-discussion segments of these documents. Create an equal number of sentences depicting overconfidence and non-overconfidence, with 10 examples of each. Overconfidence encompasses various behaviors where individuals, executives, or companies display inflated self-assessments regarding their abilities, decisions, or potential outcomes, while often underestimating the risks involved. Definitions range from general self-perceptions of superior judgment and abilities to exaggerated beliefs in one's own truth and superiority over others. This trait significantly influences decision-making processes and is reflected in risky business behaviors and aggressive corporate actions. The sentences must not contain any specific historical figures, terms, company names, events, or dates. Each sentence should be numbered (like: 1. <sentence>), and be on a new line. The output should consist only of the created list of sentences without any extra information like categorization (finance, accounting, banking, business, management, marketing, or economics) or tags (overconfidence or non-overconfidence).

Stage 2: Tagging (Temp = 0)

You are an expert with extensive experience in various fields of finance, accounting, banking, business, management, marketing, and economics. You will be provided with different sentences. Read these sentences as an expert in these fields and tag them as either 'Overconfidence', or 'Non-overconfidence'. Overconfidence encompasses various behaviors where individuals, executives, or companies display inflated self-assessments regarding their abilities, decisions, or potential outcomes, while often underestimating the risks involved. Definitions range from general self-perceptions of superior judgment and abilities to exaggerated beliefs in one's own truth and superiority over others. This trait significantly influences decision-making processes and is reflected in risky business behaviors and aggressive corporate actions. The chosen tag should reflect the specific understanding and perspective of someone from the fields of finance, accounting, banking, business, management, marketing, and economics, which could contrast with the views held by the general public. Avoid generating false or misleading tags. If you cannot generate an accurate tag, label it as 'None'. Generate only the tags as outputs with the corresponding sentence number (like: 1. <tag>). Each output should be on a new line. Input:

Real Estate Financialization in China: Theoretical Chapter

Xinyi Zhou

1. Introduction

Over the past half century, the concept of financialization has gradually entered public discourse and has been used and studied in different disciplines and fields, such as accounting and geography. Based on this, scholars have explored how financial markets, financial motives, and financial means spread across finance and beyond (Kuokkanen, 2024; Chiapello, 2015; Van der Zwan, 2014). Financialization can be defined as an accumulation mode in which profits are mainly accumulated through financial channels (Krippner, 2005). Early research conceptualized financialization primarily as a macro-level structural shift in which profits increasingly derive from financial channels rather than productive activities (Krippner, 2005; Epstein, 2005). Subsequent research expanded this perspective to the organizational level, examining how enterprises adopt a financial logic that prioritizes shareholder value, capital market expectations, and financial performance indicators (Froud *et al*, 2000; Lazonick, 2010; Van der Zwan, 2014).Focusing on financialization from the perspective of accounting research, it has mainly been explored by critical accounting researchers in the context of organizational management (Botzem and Dobusch, 2017; Froud *et al*, 2004; Graaf *et al*, 2022), accounting standard setting (Müller, 2014; Chahed, 2021), governance reforms (Modell and Yang, 2018) and sustainability (Kuokkanen, 2024; Arjaliès and Gibassier, 2023; Chua and Fiedler, 2023).

However, the large amount of existing literature on financialization is based on the experiences of developed Western economies, and implicitly assumes that financialization is mainly driven by market forces and financial markets expansion (Braga *et al*, 2017). Over the past few decades, the concept of financialization has become crucial for understanding the transformation of contemporary capitalism (Epstein, 2005: 3; Lapavitsas, 2013; Braga *et al*, 2017). Nevertheless, when examining financialization in an environment where financial markets coexist with strong state intervention, this assumption becomes problematic. China's economic activities are deeply embedded in a mixed institutional framework, in which the state power and market power are intertwined rather than mutually exclusive (Marquis and Raynard, 2015). Therefore, understanding China's financialization requires not only attention to economic dynamics, but also to the institutional logic that shapes organizational behaviour. Under this context, financialization is not solely caused by market forces; it may reflect the interaction of multiple institutional logics, which determine organizational behaviour.

By means of the institutional logic, this chapter conceptualizes the financialization process as a dynamic interaction and influence between market logic and state logic within specific institutional field. It provides a foundation for subsequent applications of the framing theory, enabling the study of how organizations interpret, argue for, and implement financialization. This chapter focuses on the real estate industry in China and constructs a theoretical framework to analyze how financialization emerged in a complex institutional environment. The real estate industry can serve as a highly significant empirical backdrop, as it has played a crucial role in driving China's economic growth while being intricately intertwined with the financial market and financial practices (Chen and Wu, 2022).

2. Financialization

2.1 Financialization Theory

Financialization is related to the process of profits accumulation in all monetary production economies and is a multi-faceted and multi-level concept (Javidanrad *et al,* 2024). Although this term is widely used across various disciplines, its meaning varies among different academic schools. The concise definition of it is even more diverse, ranging from covering "all financial activities" (Epstein, 2005) to a more narrow description of "emerging financial market activities" (Stockhammer, 2004). With the structural transformation at the macro level, the focus emphasized by the financialization theory has shifted from the dominant role of financial activities in profit creation (Krippner, 2005) to the high influence of actors in financial markets on economic policies and corporate decision-making (Epstein, 2005). Just as Ronald Dore (2008: 1097) emphasized: "'Financialization' is a bit like 'globalization'—a convenient word for a bundle of more or less discrete structural changes in the economies of the industrialized world".

Apart from the macroscopic research on financialization, another type of literature focuses on the impact of financialization at the organizational and corporate levels (Van der Zwan, 2014). This research area examines how the financial logic reshapes corporate governance, management decisions, and corporate strategies. Specifically, scholars have emphasized that the importance of shareholder value as the dominant principle in guiding corporate behaviour is increasing day by day (Rappaport, 1986). The rise of the shareholder value concept to certain extent has prompted enterprises to focus on short-term financial performance and distribute profits to shareholders rather than investing in productive capabilities (Lazonick and O'Sullivan, 2000). The empirical phenomena explored in this field include executive compensation practices, company restructurings, shareholder activism and other investor behaviors, as well as the dissemination of the shareholder value concept from the United States to other political economies (Van der Zwan, 2014). Thus, shareholder value is regarded as the "norm of capitalist transformation", and therefore provides a legitimate justification for implementing new policies and practices that are beneficial to shareholders rather than other members of the enterprise (Aglietta, 2000: 149). From this perspective, financialization can be manifested through the increasing use of financial tools, the expansion of financial investments by non-financial enterprises, and the growing influence of capital market expectations on corporate strategies.

In addition, the rest of financialization research focuses on the financialization of everyday, which stems from social accounting and cultural economics (Van der Zwan, 2014). Specifically, these studies examined projects and programs aimed at including marginalized or vulnerable groups in the financial market through participation in pension plans, housing mortgages, and other mass-market financial products (Ghafran and Yasmin, 2024; Davis and Kim, 2015; Fligstein and Goldstein, 2015;). Within its research field, individuals are encouraged to exert their new subjectivity as investors or owners of financial assets (Lai, 2017; Van der Zwan, 2014). As a result, finance becomes a decentralized form of power, which is exercised through the interaction between individuals and emerging financial technologies and financial knowledge systems (Van der Zwan, 2014).

However, with the deepening of research, researchers have gradually noticed the problems brought by financialization, such as questioning the dominant role of financial valuation, because it may reshape or even “colonize” previously non-financialized fields according to financial models (Kuokkanen, 2024; Chiapello, 2015; Golka and van der Zwan, 2022). For example, in traditionally non-financialized areas such as health care or education, financial models increasingly determine priorities and practices. In other words, these studies not only focus on how accounting financializes or contributes to financialization (Kuokkanen, 2024; Himick and Brivot, 2018), but also problematizes the increasingly important role of finance at the societal level (Graaf *et al*, 2022).

2.2 Financialization in China

Under the global trend of financialization, when placing "financialization" in the context of China, it can be found that a notable feature is that both the central and local governments in China are using financial tools to achieve policy goals (Pan *et al*., 2021; Wang, 2015). Pan *et al.* (2021) hold that China's national financialization can be identified as state-led financialization, namely, the state increasingly employs financialization policies. Moreover, Wang (2015) defined financialization in the Chinese context as the process by which the state increasingly relies on financial markets, financial indicators, and financial tools to manage assets and provide financing for public investment. This process encompasses three interrelated aspects (Wang, 2015): (a) the introduction of shareholder value orientation in asset management; (b) the expansion of non-bank financial institutions to manage related assets; (c) the provision of financing channels for state-led fixed asset investment through these institutions. These broad definitions of financialization deviate from many existing studies based on Western economies, as most of them consider financialization to be the result of neoliberalism and the weakening of state intervention (Epstein, 2005: 3; Pan *et al*., 2021; Lapavitsas, 2013; Braga *et al*., 2017). Furthermore, financial investment has always been regarded as a zero-sum game between financial activities and production activities (Wang, 2015; Boyer, 2000; Stockhammer, 2004; Crotty, 2005; Orhangazi, 2008). Due to the differences in social and institutional environments, the forms and consequences of state financialization can also vary greatly (Peck and Zhang 2013; Lapavitsas and Powell 2013). Although these two assumptions are reasonable for the situation of Western Anglo-Saxon capitalism, they do not apply universally to all economies (Wu *et al*., 2020), because the state itself is an important financial actor, especially when it attempts to link national finances with clear development concepts and national industrialization (Wang, 2015).

In the process of China's financialization, institutional factors play a crucial role (Pan *et al*., 2021). In this context, finance has not surpassed politics and replaced it, while political power can be incorporated into the explanatory framework of financialization (Wang, 2015). Given that China's economy was initially based on planning and government budgets (the socialist economy) before the Reform and Opening-Up, financialization can be regarded as a special tool added to the traditional toolkit of the state, which also includes functions such as regulation and budgeting (Wang, 2015). The state, as the dominant actor in China's economic development, has implemented and continues to undertake significant reforms to the economic and financial systems. In its policy practices, it employs various “market tools” (Wu, 2018) or “capitalist tools” (Wójcik and Camilleri, 2015) to foster economic growth and development

(Pan *et al*., 2021). During this process, the state logic is manifested in the deliberate choices of state actor and the catalytic effect of political power on financial expansion (Wang, 2015). The transformation of the governance model it brought about led to intense regional competition among local governments in implementing financialization policies (Pan *et al*., 2021), and also resulted in different policy interpretations and implementation plans. Meanwhile, due to China's national conditions and the nature of socialism (Zhao, 2009), China's financial system is largely controlled by the state. This is because most key financial institutions, including banks, securities companies, and insurance companies, are state-owned enterprises (Pan *et al*, 2021). Therefore, these state-owned financial institutions have also actively participated in the state-led financial process and hold certain advantages and policy interpretation rights.

However, a series of financial measures centred around borrowing, investment and repayment inevitably introduce cascading financial risks to political entities (Wang, 2015). At the same time, apart from private enterprises, various state institutions and state-owned enterprises' financial entrepreneurship embed market logic into political and bureaucratic procedures, thereby promoting the gradual integration and reconfiguration of the state and the market in a specific manner (Wang, 2015).

2.3 Real Estate Financialization in China

With the penetration of financialization in various fields, the concept of housing financialization is no longer unfamiliar to researchers. It is regarded as a key force driving the social and economic transformation of contemporary capitalism (Pereira, 2017). The most commonly accepted concept of housing financialization is defined as the property management, mortgage processes, and financial instruments being created or used by actors and organizations to seek profits (Sisson and Rogers, 2020). It has also brought about structural changes in the housing market, financial markets and global investment (United Nations, 2017). The financialization of real estate did not happen by coincidence, but was deeply influenced by the financialized business models (Zott and Amit, 2010), in which valuation and accounting played a fundamental role (Botzem and Dobusch, 2017). Existing relevant accounting studies have suggested that due to differences in legal systems across countries and the inherent flexibility in asset recognition and measurement, many financial services companies use valuation and accounting methods to benefit from these cross-country regulatory differences in the European real estate sector (Botzem and Dobusch, 2017). These 'boom and bust' real estate business activities promote social and economic development, but at the same time bring employment stagnation and short-termism (Lin, 2016; Botzem and Dobusch, 2017; Minsky, 1986). The multifaceted implications arising from the process of real estate financialization are not only observed in advanced western capitalist economies, but also manifest in the Chinese real estate sector. However, due to the completely different historical backgrounds and national conditions between China and these countries, these manifestations and influences also exhibit significant differences. Influenced by China's socialist nature (Zhao, 2009), the "socialist market economy" adopted by China is not equal to pure "market economies" in that it combines market mechanisms with significant state intervention and policy guidance. Rather than being governed solely by the forces of supply and demand, it depends on national policies and state control to shape economic outcomes to a significant extent (Zhao, 2009).

In the half century after China's Reform and Opening Up process started, the financialization theory has provided an important perspective for understanding the rise of China's real estate industry. The central government has gradually reduced its direct control over state-owned enterprises. This change reflects a broader process of decentralization of power and resource allocation, creating institutional space for market-oriented development (Wójcik and Camilleri, 2015). Under this institutional change, although state-owned enterprises still need to fulfil the central planning tasks, once these requirements are met, they can conduct commodity and service transactions in the market relatively independently (Wójcik and Camilleri, 2015). In this process, housing and land have been transformed from non-tradable goods allocated by the state to tradable goods that are recognized as investable assets (Chen and Wu, 2022). This laid the foundation for the emergence and development of the real estate industry and state-owned or private real estate companies. In the process of financialization, the behaviour of real estate firms is typically increasingly driven by the logic of financial markets, shareholder value, and debt financing in the context of deepening marketisation (Froud *et al*, 2000; Newberry and Robb, 2008), with Chinese real estate companies being no exception. In other words, their economic activities are increasingly oriented towards a market logic (Wu, 2015), and the accounting system is the foundational constructor of this value system (Power, 2010). However, based on state logic, even though China's national policies, land financing, and financial control providing a direction for the financialization of Chinese real estate companies, it also imposes restrictions on the unrestrained development of financialization (Chen and Wu, 2022). One example is the implementation of the "Three Red Lines" credit policy from 2021 in order to reduce the overdependence of the national economy on the real estate sector and to curb the leverage levels of real estate enterprises. It also brings potential risks and hidden dangers for China's real estate market and enterprises, such as financial fragility, liquidity risks, and policy-driven volatility.

Based on this background, real estate enterprises have gradually realized that accounting is an important means for them to avoid potential risks brought about by financialization and to legitimize their operations. Under modern accounting systems that incorporate accounting techniques such as fair value accounting, business valuations are constructed on the basis of expected future earnings and market prices, thereby transforming financial expectations based on assumptions from being recognised as unstable to being "seemingly neutral" (Power, 2010). In the context of real estate financialization, accounting can provide an "objective" finance-oriented decision-making mode through quantifiable financial expectations, thus endowing enterprises with the legitimacy of financial operations (Power, 2010; Mennicken and Power, 2015; Chiapello, 2015). Accounting can be seen as a tool to measure, present and legitimize the financial value of real estate enterprises. With the deepening of real estate financialization in China, it can be observed that the strategies and financing channels of real estate enterprises are increasingly affected by accounting indicators such as debt ratio, cash flow forecast and rate of return (Bao *et al*, 2024). Regulatory measures such as the "three Red lines" policy further highlighted how accounting data became the basis for state intervention, forcing companies to actively manage financial metrics in order to maintain credibility. Besides, the inability of the "product market" to generate sufficient returns to satisfy investor interest and maintain stock prices puts most companies under great pressure (Foster, 2007; Cushen, 2013), which is applicable to Chinese real estate enterprises as well. Faced with the possibility of divestment and hostile takeovers, and especially the loss of personal compensation that occurs to a great extent, managers are more inclined to pursue more short-term, financially short-sighted strategies to improve investor returns, such as taking on debt and limiting internal investment, even if it may harm long-term competitiveness (Cushen, 2013; Aglietta and Breton, 2001 ; Appelbaum *et al*, 2013; Lazonick, 2012). In this case, accounting provides an operational platform for their short-term financial-driven behaviours under financialization and is an important means for organizations to pursue recognition from the capital market (Mennicken and Power, 2015). It provides quantifiable and comparable financial data for executive compensation incentives, investor relations management and capital operations. However, the general observations on the role of accounting in real estate financialization mainly come from Western contexts (Botzem and Dobusch, 2017), and whether it can be fully applied to China remains an open question that requires long-term observation and further discussion.

Therefore, although the process of real estate financialization brings more opportunities and choices of financial strategies for enterprises, the pursuit of shareholder value is not a prescriptive functional strategy with predictable results (Cushen, 2013). It is described by Fraud *et al.* (2006, p.65) as a strategy that "sets management on a utopian quest for growth and high returns for capital". In fact, the consequences of such pursuits are variable and uncertain, especially in the context of China's state-controlled real estate market. In this context, understanding the financialization of China's real estate requires a perspective that goes beyond purely economic explanations, and also needs to take into account the broader institutional framework that shapes organizational behaviour. From this perspective, financialization can not only be understood as an economic transformation, but also be regarded as an institutional process embedded within a broader set of meanings, values, and practices. The institutional logic perspective provides a useful framework for analysing how actors with different interests and organizational principles shape the process of real estate financialization. In China, where the coexistence of state authority and market mechanisms is the background, financialization should be better understood as the result of the interaction of multiple institutional logics rather than merely stemming from market forces. Moreover, in China's policy-led path of real estate financialization, accounting may serve as a key tool for real estate enterprises to carry out institutional negotiations and to maintain financing channels and development space, though this requires further empirical examination.

3. Defining the Field: Real Estate Financialization in China

3.1 The Chinese Real Estate as an Issue-Based Institutional Field

In this research, the Institutional field can be seen as context or the stage for institutional changes occurred. Among the existing literature, the most widely used definition of the institutional field comes from DiMaggio and Powell (1983: 148), which is "a recognized area of institutional life: key suppliers, resource and product consumers, regulatory agencies and other organizations that produce similar services or products." Institutional fields constitute a collective of organisations sharing a common system of meaning, where interactions among participants are more frequent and significant than those with actors outside the field (Scott, 1995: 56). The field can be seen as the location of some institutions that guide their everyday behaviours (Zietsma *et al*, 2017). It represents a "local social order" (Fligstein, 2001: 107), which need to be considered by multiple actors in their actions (McAdam and Scott, 2005; Zietsma *et al*, 2017). This concept provides an important analytical starting point for understanding how the participants of an organization interact within the common field. Based on this perspective, the institutional field of China's real estate financialization is an interactive network composed of government policies, capital markets, market customers and real estate enterprises. In this network, 'financialization' is constructed, disseminated, adjusted, and reproduced with a clear guiding principle or objective (Evans and Kay, 2008: 973; Zietsma *et al*, 2017).

In the process of institutional research, how to define the boundaries of a particular field is a question that requires repeated consideration. The traditional approach usually defines institutional fields based on industry or population boundaries (DiMaggio and Powell, 1983; Scott, 2008), but this method may be overly restrictive when analyzing complex socio-economic processes across industries. The institutional field is not merely a collection of influential organizations within the same industry; indeed, it is the core of a common channel for dialogue and discussion (Hoffman, 1999). Regarding the power struggles and framing contests among members within the field (Zietsma *et al*, 2017), Hoffman (1999) argued that the institutional field is formed around a core issue, serving as the centre for the debate and interpretation negotiation where various interests interact and compete. Different from isomorphous dialogue, this process more closely resembles institutional war (White, 1992), as not all actors can influence the final debate, yet they often hold diametrically opposed views (Hoffman, 1999). These characteristics render issue-based fields competitive and dynamic, standing in clear contrast to institutional fields, which are typically regarded as possessing stable features (Wooten and Hoffman, 2008). In fact, China's real estate financialization is precisely such an issue-based field, as its development is closely linked to broader economic and political changes. This industry not only plays a crucial role in urban development and the accumulation of family wealth, but also has an important impact on fiscal governance and the expansion of the financial system. Real estate assets are increasingly being used as collateral for credit expansion, investment tools for financial institutions, and sources of fiscal revenue for local governments. At the same time, this industry remains under strong regulatory supervision and policy intervention by the central government. These overlapping functions imply that the financialization of real estate cannot merely be regarded as a transformation within the real estate industry; instead, it reflects a broader and more complex socio-economic process involving multiple parties such as finance, government, and urban governance. Moreover, it is the core issue that needs to be discussed.

4. Institutional Logics

4.1 Institutional Logics and Financialization

Friedland and Alford (1991: 241) believed that society is composed of “a set of interrelated institutional logics”. The concepts of individual, agency, freedom, and related instrumental rationality have been institutionally and historically influenced by the emergence of institutional logics such as the market, democracy, family, state, and religion (Friedland and Alford, 1991: 239-240; Quattrone, 2015).Institutional logics was defined by Thornton and Ocasio (1999: 804) as the “socially constructed, historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social reality” (Thornton and Ocasio, 1999, p. 804). Therefore, institutional logic can be regarded as the guiding principles of societies and organizations, shaping actors' perceptions while providing legitimacy and a meaning framework for their actions. Human agency and practice are in fact constrained by the rationality or logic of a series of conflicting social orders, within which diversity in organizational fields emerges (Friedland and Alford, 1991). Financialization not only involves the expansion of financial activities, but also encompasses deeper-level changes at various levels such as corporate governance, accounting standard setting, sustainability, investment focus, and economic policies (Botzem and Dobusch, 2017; Froud *et al*, 2004; Müller, 2014; Modell and Yang, 2018; Kuokkanen, 2024; Chua and Fiedler, 2023). From this perspective, financialization is not merely an economic process, but also an institutional transformation. It has altered the way participants understand value creation, risks, and economic growth.

Although most of the literature on financialization has focused more on the structural economic changes (Dávila-Fernández and Punzo, 2018; Botzem and Dobusch, 2017), the institutional logic perspective emphasizes the cultural and ideological foundations that facilitate such changes (Thornton *et al*., 2004). Early empirical studies of institutional logics emphasized the temporal changes of dominant logics, namely, the collapse and replacement of mainstream logics, as well as their consequences for organizations and fields (Purdy and Gary, 2009; Haveman & Rao 1997, Lounsbury 2002, Thornton & Ocasio 1999). They believed that the central logic would control the operation of the extensive departmental activity system, and social actions are formed in the conflicts and tensions between and within these interrelated networks (Quattrone, 2015). In other words, this is an effective mapping of the logical explanatory ability (Lounsbury *et al*, 2021). Thus, logic is increasingly being used to explain the historical contingency of institutions and the dynamics within the organizational field (Greenwood *et al*., 2011). Adopting the institutional logic to explain the process of financialization is logical and effective. At the same time, financialization can be understood as the increasing dominance of a market-oriented logic, which regards shareholder value, capital accumulation and financial performance as the key indicators for measuring the success of an enterprise (Epstein, 2005; Rappaport, 1986; Van der Zwan, 2014). Under this logic, enterprises are increasingly adopting strategies that are oriented towards financial returns, such as leveraging, asset transactions and financial investments.

However, the process of financialization is rarely determined solely by market forces; instead, it is the result of the combined influence of financial markets, financial actors, and financial motives (Epstein, 2005). Based on different background, financialization unfolds through the interaction between market logic and other institutional orders. This is in line with the research results of Purdy and Gary (2009), the standard diffusion model of institutions (with a single dominant logic) may not be sufficient to explain how new organizational groups establish their positions in emerging institutional fields. In the changing institutional field, there can exist multiple competing organizational logics, which shape actions in different ways. Institutional logics are the fundamental frameworks with cultural cognitive significance that constitute actions, and institutional change occurs when the structure, dominance, or interaction of these logics' changes (Thornton *et al*., 2004). Therefore, the financialization process often reflects the negotiation and interaction between different institutional logics, rather than merely the expansion of one aspect of the financial market. By viewing financialization through the lens of institutional logic, the analytical focus shifts to the beliefs, interests, priorities and strategic directions of the key participants. Different participants may interpret the meaning and purpose of financialization in different ways based on the institutional logic that guides their actions. The different solutions proposed by multiple actors from different standpoints originate from the institutional logics that already exist or are developing in the broader social domain, especially in the context of major social debates related to economic and livelihood issues (Jonsson and Lounsbury, 2017). It is common for organizations to encounter multiple logics, and these logics may be incompatible with each other (Friedland & Alford, 1991; Kraatz &

Block, 2008; Selznick, 1949). Therefore, in a specific industry or national context, the evolution of financialization is determined by the competing logics and the interactions among the participants who mobilize these logics.

4.2 Institutional Logics in the real estate financialization in China

Following the Reform and Opening-up initiated in 1978, China underwent a broad transformation at the level of the national political–economic system, shifting from a centrally planned socialist economy to what has been described as a “socialist economy with Chinese characteristics” (Zhao, 2009). This economic model is a combination of socialist economy and market economy. Specifically, during China's economic transformation, market mechanisms have gradually been introduced into the political system, which is still profoundly influenced by state authority. In other words, it is a market economy under the leadership of the government. Compared to other actors, the state possesses an absolutely powerful capability to mobilize political and financial resources and to utilize and promote financial tools in order to achieve development goals (Pan *et al*, 2020). As a crucial player in the economic growth, the Chinese government has been "actively advancing the growth agenda" by continuously proposing and adjusting policies (Wu, 2018; Pan *et al*., 2020). Studies related to China's financialization indicate that, despite its market-oriented transformation, all financialization participants are often deeply embedded in the political structure and the state-led development strategies (Wu, 2002; Pan *et al*, 2020; Wang, 2015;). In this institutional environment, the state continues to play a core role in resource allocation, policy focus setting, and economic development guidance, while market activities operate within the framework of strong state influence. The government may actively promote financialization as part of a broader economic development strategy, while also striving to regulate or control its consequences (Chen and Wu, 2022). Therefore, in the China’ s real estate financialization, the interaction between the state logic and the market logic is particularly evident. Rather than directly determining developments in specific sectors, these reforms reconfigured the broader institutional environment and dominant logics within which more specific issue-based fields, such as real estate, have subsequently emerged and evolved.

China’s real estate sector can thus be seen as a nested sub-field that is embedded within and shaped by, but not reducible to, these wider broader political transformations. Based on this perspective, in the field of China’s real estate financialization, multiple logics such as the state logic and market logic coexist, as conceptualised within Thornton *et al***.**’s (2012: 17) interinstitutional system. Following Thornton *et al* (2012: 17), state and market logics are understood as ideal-typical institutional orders, each characterised by distinct organizing principles, sources of legitimacy, and authority structures. For the state logic within its field, the financialization of China's real estate implies stable growth of China's macroeconomy, social stability, and local land fiscal revenue; while the market logic emphasizes supply and demand relationship, profit, risk management and investment return. It demands the emergence of leverage, capital operation and financial productization in the process of China's real estate financialization to achieve high returns. Although these logics exist in contradiction to each other, they are not mutually exclusive, but interdependent (Jonsson and Lounsbury, 2017). Based on the institutional environment, various actors often need to coordinate and handle situations where multiple logics coexist. On one hand, the market-oriented incentive mechanism prompts actors to pursue profits, expansion, and capital accumulation. On the other hand, political priorities such as economic stability, social development, and strategic industries remain the core content of national governance. For instance, during the continuous reconfiguration of the real estate financialization field, real estate enterprises, influenced by market logic, attempt to expand and obtain more profits through an operation system featuring high leverage and high debt. Meanwhile, under the dominance of the state logic, the government proposes restrictive policies (such as the Three Red Lines) and improves regulations to curb the excessive development of real estate financialization and the associated financial risks. However, from another perspective, the development of real estate enterprises requires the support of national policies. The real estate industry serves as the largest reservoir of capital, and the Chinese economy to a certain extent relies on the vigorous development of the real estate sector. Moreover, due to China's fiscal policies, local governments rely on land sales for revenue. This interdependent relationship to some extent helps to explain why China's economy has long been dependent on finance and land.

The coexistence of this national logic and market logic is of great significance for understanding the phenomenon of real estate financialization in China. Financial expansion may be driven by market incentives such as profit opportunities and investment demands, but it may also be influenced by national-led development strategies and policy goals. Therefore, when making strategic decisions, actors must balance these expectations that are in conflict with each other. Due to the complexity of the institutional environment and logic, the development of China's real estate financialization cannot be explained merely by market dynamics; instead, it must be regarded as the result of the interaction among multiple institutional logics and the joint actions of the actors involved (Purdy *et al*, 2019).

4.3 Critics on Logics and Supplement

Logic is often regarded as a set of stable beliefs and assumptions, and the differences among logics are the root cause of practical variations (Lounsbury, 2008). In the studies related to institutional logics, most of them assumed that institutional logics are well established, stable, or exists in the same way as “social facts” (Purdy *et al,* 2019). People do not believe that the vitality of logic lies within logic itself. Broader institutional vitality stems from the competition between the logic of competition and institutional actors (Reay and Hinings, 2009), from tensions and institutional changes (Thornton and Ocasio, 1998; Seo and Creed, 2002), or from the agency of institutional entrepreneurs (DiMaggio, 1988; Garud *et al*, 2007). Based on this excessive static view of meaning making, it usually serves to explain the actions of actors, but fails to explain the process of institutional construction. Purdy *et al.* (2019: 1) emphasized such research “does not attend to the interactions and processes through which meanings and practices are not just used or recombined but also initiated, reconstituted, or instantiated at multiple levels of social organization”.

Furthermore, many studies on institutional logic have been guided by a certain epistemological perspective (Modell, 2026). This epistemology neglects or downplays the complex dynamics through which institutional logic can be implemented, replicated and transformed, and tends to idealize social order as an ideal type or analytical prototype (Alvesson *et al*., 2019; Alvesson and Spicer, 2019; Jackson *et al*., 2019). In the process of applying the institutional logic, researchers often simplify or generalize complex fields into several fixed logics, believing that they are replicated in a “top-down” manner at different levels of analysis. However, the actual significance of these logics is often more diverse, ambiguous and variable (Currie and Spyridonidis, 2016). This perspective of institutional logic may lead to the disconnection of organizational research from its actual context, but it has gradually become a dominant research method (Alvesson *et al*., 2019; Alvesson and Spicer, 2019; Jackson *et al*., 2019). Some studies on institutional logic have overly focused on individual institutional fields, ignoring the perspective of Friedland and Alford (1991: 241), which is all about “bringing society back in” institutional theory. In other words, they have not been deeply rooted in the organizational domain and its daily practices. Moreover, the researchers emphasized that when the institutional logic perspective is applied to situations beyond those originally targeted by the

Western capitalist society, the trend of decontextualization becomes necessary and challenging (Modell, 2026; Purdy *et al*., 2019). They argued that when the identified order is extended to other contexts, a high degree of caution is necessary (Purdy *et al*., 2019). More radically or critically, the institutional logic perspective might need to be replaced by more open research methods (Purdy *et al*., 2019).

Solving this problem shifts the focus of the analysis to the process of frames construction. This chapter supports the part of viewpoints of Purdy *et al*. (2019) and holds that the framing theory serves as an important supplementary tool for understanding the formation of institutional meanings. The framing mechanism enables participants to selectively apply, reinterpret and integrate various institutional logics, thereby providing a legitimacy basis for specific action plans (Cornelissen *et al*., 2015, p. 14). By studying how framing expresses and debates various logics, the structural diversity can be linked to the dynamic process of institutional change. Therefore, by integrating the perspective of institutional logic with the framing theory, it is possible to provide a more comprehensive explanation of how the tensions within the institutional field of real estate financialization in China are manifested and restructured over time.

5. Framing and Framing contests in Institutional change

5.1 Framing in Institutional theory

Frames define the interpretive schema that people use to locate and understand world events (Goffman, 1974: 21). Framing can be conceptualized as an active and strategic process where participants are able to define the problem, assign responsibility, and propose solutions (Benford and Snow, 2000). Framing theory offers a recursive perspective, suggesting that institutions are generated and reproduced through the daily activities of individuals (Purdy *et al*., 2019). The existing interpretive models provided by frames and the interactive process of meaning making offered by framing are ideal tools for studying and explaining the communicative composition, maintenance and transformation of institutions (Cornelissen *et al*., 2015, p. 14). It is in line with the social constructivism aspect of institutional theory (Gray *et al*., 2015). It is a constitutive process through which reality is constructed within society. By defining which issues are regarded as reasonable ones, goals and actions, the frames shape the scope of institutional possibilities. In the institutional theory, the concept of framing is employed to analyze how actors construct legitimacy and shape discourse at the field level.

The maintenance of an institution is not only dependent on formal rules and material practices, but also on the discursive process that can stabilize shared meanings (Phillips *et al*., 2004; Hardy *et al*., 1999). Therefore, the frames links discourse to institutionalization: through the framing, specific interpretations are shaped, while other interpretations are marginalized.

The framing construction of institutional theory possesses two distinct advantages, including capturing the institutionalized persistent meaning structures, and providing a macro-level structural foundation for the motivations, cognition and discourse of micro-level actors (Cornelissen and Werner, 2014: 29-30). It not only exists a priori, but also involves active struggles and negotiations over meaning before it becomes solidified and institutionalized, thereby triggering dynamic processes of meaning construction within and across groups, organizations and fields (Gray *et al*, 2015). In some of these studies related to meaning making methods and field logics, some perspectives have tended to emphasise a top-down approach, i.e. how field logics shape frames, schemas and narratives (Thornton *et al.,* 2012). Regarding how such meanings are constructed and transformed through micro-processes, the insights they provide are limited. It can also stem from bottom-up processes that continually aggregate and "amplify" existing logics, thereby challenging and reshaping them (Purdy *et al,* 2019). However, framing actually has a bidirectional nature and interactivity (Gray *et al,* 2015: 116).

5.2 Framing Contests and Logics

The frames provide contextual and interactive micro-foundations for the implementation and potential transformation of the logic, but the frames themselves are constrained by the institutional logic embedded in a particular social context. Some theoretical models provide an overview of how the explanatory process influences strategies (Kiesler and Sproull 1982, Ocasio 1997), and some empirical studies have shown that managers' frames shape the interpretation of the environment during times of turbulence and subsequent strategic choices (Barr 1998, Barr *et al.* 1992, Tripsas and Gavetti, 2000). However, due to the presence of numerous participants with diverse interests and identities in the institutional field, the process of framing construction rarely proceeds without any disputes. Multiple cognitive frames can coexist in the same field, so any action model in these contexts needs to consider how certain frames are superior to others (Kaplan, 2008). The field has transformed into a battleground for framing construction, with ongoing struggles over how to define the problem, which values should hold the dominant position, and which action plans should be regarded as legitimate. The existence of this tense relationship-induced framing contests has been proven to be the central mechanism for institution establishment (Kaplan, 2008; Schneiberg & Soule, 2005; Guerard *et al*., 2013). Therefore, the concept of framing contests proposed by Kaplan (2008) needs to be discussed, which explaining how the actors attempt to transform their individual cognitive frames into the main collective frame of the institution through their daily interactions. Actors often engage in framing practices in an attempt to resonate their decision-making frames and mobilize action in their favour, ultimately establishing or recalibrating dominant frames to influence the positions of others' strategic choices and legitimize their own claims and interests (Kaplan, 2008). For example, in the institutional field of real estate financialization in China, the central government proposed frames such as "housing for living purposes and not for speculation" and "ensuring financial stability", and then strengthened financial supervision to require real estate enterprises to deleverage and control risks; influenced by China's fiscal and taxation policies and the performance evaluation criteria of early officials which were based on GDP growth and fiscal revenue, local governments implemented the policy of "maintaining growth" and relied on land finance; during this period of expected high demand for real estate, real estate enterprises insisted on frames such as "the market has demand" and "loose financing" in order to maximize profits. The long-term competition among these different frames eventually led to the formation of the Chinese-style trajectory of real estate financialization.

However, it is obvious that the unique characteristics of China's real estate financialization field have led to an asymmetric framing contest. When a frame resonates with broader cultural narratives, its influence grows as it demonstrates a high degree of alignment with the beliefs, values, and experiences of potential adherents (Snow and Benford, 1988). This is in line with the notion of social situations and situational frames emphasized by Furnari (2019). Framing occurs within a specific context, which is jointly shaped by the field structure, power relations, existing institutional logic, and material and symbolic resources (Furnari, 2019). When there is a relatively stronger decision-making frame in place, other actors regard it as a constraint as well as a resource. In order to legitimize their actions and decision-making frames, they find an intermediate position between the cognitive model and the political model for strategic formulation through purposeful interaction (Kaplan, 2008). In the context of the real estate financialization in China, actors other than the state would recognize that the current situation is a typical scenario within the state's institutional order, consisting of certain roles and character frames (Furnari, 2019). Generally speaking, they adopt an image of absolute obedience or compliance to play their roles. Specifically, in order to rationalize their profit-seeking behaviour within the state logic, Chinese real estate enterprises in this period placed themselves in the role of nation-builders or supporters of urbanization through framing. Instead of describing pure reality, they are constructing actionable, purposeful realities that provide legitimacy for strategic actions when framing, such as financing, expansion, transformation.

In summary, the framing contests reveals the political aspect of institutional change. Different frames do not simply coexist; instead, they compete with each other to establish dominance. When a certain framework gains widespread acceptance, it can restructure the hierarchy of legitimacy, redefine the governance mechanism, and change the evaluation standards in that field. From another perspective, continuous contests may lead to hybrid arrangements that incorporate elements of multiple logics.

6. Integrating Field, Logics and Framing

In condition, in the sections above, the institutional field is described as an interactive space with structural characteristics (DiMaggio and Powell, 1983), the institutional logic is regarded as the diverse cultural order that organizes these arenas (Jonsson and Lounsbury, 2017), and framing as the interpretive mechanism through which actors mobilize and contest meanings (Benford and Snow, 2000). However, considered independently, each perspective can only offer a partial explanation of institutional change. Therefore, it is necessary to establish a comprehensive framework to explain how structural conditions, cultural diversity, and discourse agency interact with each other during the process of field-level changes.

The institutional fields constitute the structural environment in which actors operate, determining the way resources are allocated and the stability of governance arrangements (Oakes *et al*, 1998: 260). Within these fields, multiple institutional logics coexist, each providing competing normative principles and evaluation criteria (Friedland & Alford, 1991). This diversity leads to ambiguity in interpretation (Currie and Spyridonidis, 2016) and disconnection between the organization of research and the actual background (Alvesson *et al*., 2019; Alvesson and Spicer, 2019; Jackson *et al*., 2019), ignoring the "bringing society back in" institutional theory (Friedland & Alford, 1991). However, these logics do not automatically translate into actions. They must be activated, clarified, and applied in a strategic manner. Framing serves as the mediating mechanism linking logics to institutional outcomes. Actors draw selectively upon available logics to construct problem definitions, justify policy directions, and claim legitimacy. Through the framing processes, specific logics are highlighted, while others are marginalized or reinterpreted (Phillips *et al*., 2004; Hardy *et al*., 1999). When competing actors propose alternative frames, framing contests emerges, thereby transforming the field into an arena for a discourse struggle (Kaplan, 2008). The outcomes of these conflicts can reshape the hierarchical structure of legitimacy, reconfigure the relative dominance of logic, and ultimately alter the field structure.

In details, the chapter conceptualize China's real estate field as a configuration shaped by multiple institutional logics, including state logic and market logic, and as a process of coordinating and stabilising these logics through a series of framing practices (Jonsson and Lounsbury, 2017). In this institutional field, the significance of China's real estate financialization has evolved continuously through ongoing competitive framing contests (Kaplan, 2008). During this process, the state logic plays a certain role in determining the emergence of the frames, while the actors selectively emphasizing or reinterpreting certain logics to achieve their own interests (Kaplan, 2008). This chapter does not view the institutional logic as a fixed background condition, nor framing as unconstrained agency. This comprehensive perspective conceptualizes institutional change as a dynamic process: the field structure determines the availability and authority of the logic; the logic provides cultural materials for framing; and framing contests reshape the configuration of the logic and the structure of the field. By tracing these interrelated processes, the framework constructed here connects structural and discursive institutionalism and specifies the micro–macro linkage through which institutional pluralism generates transformation.

Reference

Aglietta M. (2000), Shareholder Value and Corporate Governance: Some Tricky Questions, *Economy and Society*, 29, pg. 146-159

Aglietta, M. and Breton, R. (2001) 'Financial systems, corporate control and capital accumulation', *Economy and Society*, 30 (4), pp. 433-466

Alvesson, M. & Spicer, A. (2019). Neo-institutional theory and organization studies: a mid-life crisis?, *Organization Studies*, 40(2), 199-218.

Alvesson, M., Hallett, T. & Spicer, A. (2019). Uninhibited institutionalisms. *Journal of Management Inquiry*, 28(2), 119-127.

Appelbaum, E. and Batt, R. and Clark, I. (2013) 'Implications of financial capitalism for employment relations research: Evidence from breach of trust and implicit contracts in private equity buyouts', *British Journal of Industrial Relations.*

Arjaliès, D.-L. and Gibassier, D. (2023) 'Can Financialization Save Nature? The Case of Endangered Species.' *Contemporary Accounting Research*, 40, pp. 488-525. https://doi-org.bris.idm.oclc.org/10.1111/1911-3846.12810

Bao, H and Wang, Z and Wu, R (2024) 'Understanding local government debt financing of infrastructure projects in China: Evidence based on accounting data from local government financing vehicles', *Land Use Policy*, 136, 106964, ISSN 0264-8377, https://doi.org/10.1016/j.landusepol.2023.106964.

Barr, P. S. (1998). Adapting to unfamiliar environmental events: A look at the evolution of interpretation and its role in strategic change. *Organization Science*, 9(6) 644–669.

Barr, P. S., J. L. Stimpert, A. S. Huff. (1992). Cognitive change, strategic action, and organizational renewal. *Strategic Management J.* 13(Special Issue) 15–36.

Battilana J (2006) Agency and institutions: The enabling role of individuals' social position. *Organization* 13(5): 653–676.

Benford, Robert & Snow, David. (2000). Framing Processes and Social Movements: An Overview and Assessment. *Annual Review of Sociology*. 26. 611-639. 10.1146/annurev.soc.26.1.611.

BESHAROV, M. L., & SMITH, W. K. (2014). MULTIPLE INSTITUTIONAL LOGICS IN ORGANIZATIONS: EXPLAINING THEIR VARIED NATURE AND IMPLICATIONS. *The Academy of Management Review*, 39(3), 364–381. http://www.jstor.org/stable/43699249

Botzem, S and Dobusch, L. (2017) 'Financialization as strategy: Accounting for inter-organizational value creation in the European real estate industry', *Accounting, Organizations and Society,* 59, pp. 31-43

Braga, J., Oliveira, G. C. D., Wolf, P. J. W., Palludeto, A. W. A., & Deos, S. S. D. (2017). For a political economy of financialization: Theory and evidence. *Economia e Sociedade*, 26(Spe), 829–856. doi: https://doi.org/10.1590/1982-3533.2017v26n4art1.

Brint, S., & Karabel, J. 1991. Institutional origins and transformations: The case of American community colleges. In W. W. Powell & P. J.DiMaggio (Eds.), *The new institutionalism in organizational analysis*: 337-360. Chicago: University of Chicago Press.

Chahed, Y. (2021), 'Words and Numbers: Financialization and Accounting Standard Setting in the United Kingdom.' Contemporary Accounting Research, 38, pp.302-337.

Chen, J and Wu, F (2022) 'Housing and land financialization under the state ownership of land in China', Land Use Policy, Volume 112, 104844, ISSN 0264-8377, https://doi.org/10.1016/j.landusepol.2020.104844.

Chiapello, E (2015) 'Financialisation of valuation', Human Studies, 38 (1), pp. 13-35

Chua, W.F. and Fiedler, T. (2023) 'Engaging with matters that matter: A cross-cultural journey into the financialization of climate change effects', Qualitative Research in Accounting & Management, 20 (1), pp. 92-116

Cornelissen, J. P., & Werner, M. D. (2014). Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. *Academy of Management Annals*, 8: 181–235.

Cornelissen, J. P., Durand, R., Fiss, P. C., Lammers, J. C. and Vaara, E. (2015). Putting communication front and center in institutional theory and analysis. *Academy of Management Review,* 40, 10– 27.

Cornelissen, J., Mantere, S., & Vaara, E. (2014). The contraction of meaning: The combined effect of communication, emotions, and materiality on sensemaking in the Stockwell shooting. *Journal of Management Studies*, 51: 699 –736.

Currie, G., & Spyridonidis, D. (2015). Interpretation of Multiple Institutional Logics on the Ground: Actors' Position, their Agency and Situational Constraints in Professionalized Contexts. *Organization Studies*, 37(1), 77-97. https://doiorg.manchester.idm.oclc.org/10.1177/0170840615604503 (Original work published 2016)

Cushen, J. (2013) 'Financialization in the workplace: Hegemonic narratives, performative interventions and the angry knowledge worker', *Accounting, Organizations and Society*, 38 (4), pp. 314-331, ISSN 0361-3682,

David A. Snow & Robert D. Benford (1988) Ideology, frame resonance, and participant mobilization. *International Social Movement Research, 1, 197–217.*

Dávila-Fernández, M. J., & Punzo, L. F. (2020). Financialisation as structural change: measuring the financial content of *things*. *Economic Systems Research*, *32*(1), 98–120. https://doi.org/10.1080/09535314.2019.1643294

Davis, G., & Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: Institutional fields and mechanisms. *Organization Science*, 16: 332-343.

Davis, K and Kim, S. (2015) Financialization of the Economy, Annual Review of Sociology, 41, pp. 203-221

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review,* 48(2), 147–160. https://doi.org/10.2307/2095101

DiMaggio, Paul J., and Walter W. Powell. (1991) 'Introduction.' Pp. 1–40 in The New Institutionalism in Organizational Analysis, edited by Paul J. DiMaggio and Walter W. Powell. Chicago: The University of Chicago Press.

Dore R. (2008), Financialization of the Global Economy, Industrial and Corporate Change, 17 pg. 1097-1112

Eitrem, A., Meidell, A. and Modell, S. (2024) ‘The use of institutional theory in social and environmental accounting research: a critical review’, *Accounting and Business Research*, 54(7), pp. 775–810. doi: 10.1080/00014788.2024.2328934.

Epstein, G. (2005) *Financialization and the world economy*, Edward Elgar, Cheltenham, pp. 3-17

Evans, R. and Kay, T. (2008). ‘How environmentalists “greened” trade policy: Strategic action and the architecture of field overlap.’, *American Sociological Review*, 73, pp.970–991.

Fligstein, N and Goldstein, A. (2015) The emergence of a finance culture in American households, 1989–2007, *Socio-Economic Review*, 13 (3), pp. 575-601

Fligstein, N. (2001) ‘Social skill and the theory of fields.’, *Sociological Theory*, 19, pp.105–125.

Fligstein, N., & McAdam, D. (2012). *A theory of fields*. Oxford, UK: Oxford University Press.

Fligstein, N (1990) *The transformation of corporate control*. Cambridge, MA: Harvard University Press.

Foster, J. B. (2007). ‘The financialization of capitalism.’, *Monthly Review, 58*, 11.

Friedland R. and Alford R. (1991) ‘Bringing society back in: Symbols, practices, and institutional contradictions.’ In Powell W. W., DiMaggio P. J. (eds.), *The New Institutionalism in Organizational Analysis*, pp.232–266. Chicago: University of Chicago Press.

Froud, J. and Haslam, C. and Johal, S. and Williams, K. (2000) ‘Shareholder value and financialization: consultancy promises, management moves, Economy and Society, 29 (1), pp. 80-110

Froud, J. and Johal, S. and Papazian, V. and Williams, K. (2004) ‘The temptation of Houston: A case study of financialisation’, Critical Perspectives on Accounting, 15 (6-7), pp. 885-909

Furnari, S. (2016). Institutional fields as linked arenas: Inter-field resource dependence, institutional work and institutional change. *Human Relations*, 69(3), 551-580. https://doi-org.manchester.idm.oclc.org/10.1177/0018726715605555 (Original work published 2016)

Furnari, S. (2019). "Situating Frames and Institutional Logics: The Social Situation as a Key Institutional MicroFoundation", *Microfoundations of Institutions*, Patrick Haack, Jost Sieweke, Lauri Wessel

Garud, R., C. Hardy, and S. Maguire (2007) "Institutional entrepreneurship as embedded agency: An introduction to the special issue*." Organization Studies*, 28: 957-969

Ghafran, C and Yasmin, S. (2024) Beyond conventional financialization: Intersectional insights and Indigenous responses to financial inequality in the UK, Critical Perspectives on Accounting, 100, 102771,ISSN 1045-2354, https://doi.org/10.1016/j.cpa.2024.102771.

Giddens, A. (1979) *Central problems in social theory: Action, structure, and contradiction in social analysis*, Vol. 241, Univ of California Press

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. Cambridge, MA: Harvard University Press.

Golka, P. and Van der Zwan, N. (2022) 'Experts versus representatives? Financialised valuation and institutional change in financial governance', New Political Economy, pp. 1-14

Graaf, J. and Kraus, K. and Strömsten, T. (2022) 'The problematics of financialization–On the important (but neglected) horizontal axis of organizational action', Critical Perspectives on Accounting, 89.

Gray B., Purdy J. M., Ansari S. (2015). 'From interactions to institutions: Microprocesses of framing and mechanisms for the structuring of institutional fields.' *Academy of Management Review*, 40, pp.115-143.

GRAY, B., PURDY, J. M., & ANSARI, S. S. (2015). FROM INTERACTIONS TO INSTITUTIONS: MICROPROCESSES OF FRAMING AND MECHANISMS FOR THE STRUCTURING OF INSTITUTIONAL FIELDS. The Academy of Management Review, 40(1), 115–143. http://www.jstor.org/stable/43700545

Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. Organizational Research Methods, 13: 620–643.

Greenwood, R. and Raynard, M and Kodeih, F and Micelotta, E and Lounsbury, M (2011) Institutional Complexity and Organizational Responses. ANNALS, 5, 317–371, https://doi-org.manchester.idm.oclc.org/10.5465/19416520.2011.590299

Hardy, C., & Phillips, N. (1999). No Joking Matter: Discursive Struggle in the Canadian Refugee System. Organization Studies, 20(1), 1-24. https://doi-org.manchester.idm.oclc.org/10.1177/0170840699201001 (Original work published 1999)

Hargrave, T., & Van de Ven, A. (2006). A collective action model of institutional innovation. Academy of Management Review, 31: 864-888.

Haveman HA, Rao H (1997) Structuring a theory of moral sentiments: Institutional and organizational coevolution in the early thrift industry. American Journal of Sociology 102(6): 1606–1651.

Haveman HA, Russo MV, Meyer AD (2001) Organizational environments in flux: The impact of regulatory punctuations on organizational domains, CEO succession, and performance. Organization Science 12(3): 253–273.

Hensmans, Manuel (2003) 'Social movement organizations: A metaphor for strategic actors in institutional fields'. Organization Studies, 24/3: 355–381.

Himick, D. and Brivot, M. (2018) 'Carriers of ideas in accounting standard-setting and financialization: The role of epistemic communities', Accounting, Organizations and Society, 66, pp. 29-44

Hoffman, A. J. (1999). Institutional evolution and change: Environmentalism and the U.S. chemical industry. Academy of Management Journal, 42(4), 351-371. Retrieved from https://manchester.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/institutional-evolution-change-environmentalism-u/docview/199773055/se-2

Jackson, G., Helfen, M., Kaplan, R. Kirsch, A. & Lohmeyer, N. (2019). The problem of de-contextualization in organization and management research. Research in the Sociology of Organizations, 59, 21-42.

Javidanrad, F.A. and Ackrill, R and Bakas, D & Garratt, D.(2024) ‘Theorizing the process of financialization through the paradox of profit: The credit-debt reproduction mechanism.’ Journal of Post Keynesian Economics, 47, pp.566–588.

Jonsson, S and Lounsbury, M. (2016) ‘The Meaning of Economic Democracy: Institutional Logics, Parabiosis, and the Construction of Frames’, How Institutions Matter.

Kaplan, S. (2008) ‘Framing Contests: Strategy Making Under Uncertainty’, Organization Science, 19 (5), pp. 729-752

Kiesler, S., L. Sproull. (1982). Managerial response to changing environments: Perspectives on problem sensing from social cognition. Administrative Science Quarterly. 27(4) 548–570.

Kraatz, M.S., & Block, E.S. (2008). Organizational implications of institutional pluralism. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), The SAGE handbook of organizational institutionalism (pp. 243–275). London: Sage.

Krippner, G (2005) ‘The financialization of the American economy’, *Socio-Economic Review*, 3 (2), pp. 173–208, https://doi.org/10.1093/SER/mwi008

Kuokkanen, N. (2024) ‘A problematizing review of the financialization of living beings’, Critical Perspectives on Accounting, 99, 102739, ISSN 1045-2354, https://doi.org/10.1016/j.cpa.2024.102739.

Lai, K.P (2017) Unpacking financial subjectivities: Intimacies, governance and socioeconomic practices in financialisation, Environment and Planning D, 35 (5), pp. 913-932

LANGLEY, A., SMALLMAN, C., TSOUKAS, H., and VAN DE VEN, A. H. (2013). ‘PROCESS STUDIES OF CHANGE IN ORGANIZATION AND MANAGEMENT: UNVEILING TEMPORALITY, ACTIVITY, AND FLOW.’ The Academy of Management Journal, 56(1), pp.1–13.

Lapavitsas, C, and Powell, J. (2013). “Financialization varied: a comparative analysis of advanced economies.” Cambridge Journal of Regions, Economy and Society 6(3), 359–379.

Lapavitsas, C. (2013). “The Financialization of Capitalism: ‘Profiting without Producing.” City 17 (6): 792–805. https://doi.org/10.1080/13604813.2013.853865.

Lazonick, W. (2010) ‘Innovative Business Models and Varieties of Capitalism: Financialization of the U.S. Corporation’, Business History Review, 84(4), pp. 675–702. doi:10.1017/S0007680500001987.

Lazonick, W. (2012). *The financialization of the US Corporation: What has been lost, and how it can be regained*. MPRA paper no. 72307.

Lazonick, W., & O’Sullivan, M. (2000). Maximizing shareholder value: a new ideology for corporate governance. Economy and Society, 29(1), 13–35. https://doi.org/10.1080/030851400360541

Leblebici H, Salancik GR, Copay A, King T (1991) Institutional change and the transformation of interorganizational fields: An organizational history of the U.S. radio broadcasting industry. Administrative Science Quarterly 36(3): 333–363.

Leibel, E., Hallett, T., & Bechky, B.A. (2017). Meaning at the Source: The Dynamics of Field Formation in Institutional Research. The Academy of Management Annals, 12, 154-177.

Lin, K. (2016) ‘The rise of finance and firm employment dynamics’, Organization Science, 10.1287/orsc.2016.1073

Lounsbury, M (2008) Institutional rationality and practice variation: New directions in the institutional analysis of practice, Accounting, Organizations and Society, 33, Issues 4–5, pp. 349-361

Lounsbury, M & Steele, C & Wang, M & Toubiana, M. (2021). New Directions in the Study of Institutional Logics: From Tools to Phenomena. Annual Review of Sociology. 47. 10.1146/annurev-soc-090320-111734.

Lounsbury, M. (2002). Institutional transformation and status mobility: The professionalization of the field of finance. Academy of Management Journal, 45: 255-266.

Maguire S, Hardy C, Lawrence TB (2004) Institutional entrepreneurship in emerging fields: HIV/AIDS treatment advocacy in Canada. Academy of Management Journal 47(5): 657–679.

Marquis, C and Raynard, M. (2015). Institutional Strategies in Emerging Markets. Academy of Management Annals. 9. 291-335. 10.5465/19416520.2015.1014661.

McAdam, D., and Scott, W. R. (2005) 'Organizations and movements.' Social movements and organization theory, pp. 4–40. Cambridge, UK: Cambridge University Press.

Mennicken, A., and Power, M. (2015). 'Accounting and the plasticity of valuation.' In A. B. Antal, M. Hutter, & D. Stark (Eds.), Moments of valuation: Exploring sites of dissonance (pp. 208–228). Oxford University Press.

Meyer AD (1982) Adapting to environmental jolts. Administrative Science Quarterly 27(4): 515–537.

Meyer, J., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. American Journal of Sociology, 83, 340-363.
http://dx.doi.org/10.1086/226550

Minsky, H.P. (1986) Stabilizing an unstable economy, Mcgraw-Hill, New York

Modell, S and Yang, C (2018) 'Financialisation as a strategic action field: An historically informed field study of governance reforms in Chinese state-owned enterprises', Critical Perspectives on Accounting, 54, pp.41-59

Modell, S. (2026). Institutional Logics versus Norm Circles: Alternative or Complementary Approaches to the Analysis of Institutions?

Müller, J. (2014) 'An accounting revolution? The financialisation of standard setting', Critical Perspectives on Accounting, 25 (7), pp. 539-557

Newberry, S. and Robb, A. (2008) 'Financialisation: Constructing shareholder value… for some', Critical Perspectives on Accounting, 19, pp. 741-763.

Oakes, L. S., Townley, B., and Cooper, D. J. (1998) ‘Business planning as pedagogy: Language and control in a changing institutional field.’, Administrative Science Quarterly, 43, pp.257–292.

Ocasio, W. (1997) Towards an attention-based view of the firm. Strategic Management J. 18 187–206.

Pan, F., Zhang, F., & Wu, F. (2021). State-led financialization in china: The case of the government-guided investment fund. The China Quarterly, 247, 749-772. doi:https://doi.org/10.1017/S0305741020000880

Pereira, A.L.D.S. (2017), ‘Financialization of Housing in Brazil: New Frontiers.’, International Journal of Urban and Regional Research, 41, pp.604-622.

Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and Institutions. The Academy of Management Review, 29(4), 635–652. https://doi.org/10.2307/20159075

Power, M. (2010). ‘Fair value accounting, financial economics and the transformation of reliability.’, Accounting and Business Research, 40(3), pp.197–210.

Purdy, J and Gray, B. (2009) Conflicting Logics, Mechanisms of Diffusion, and Multilevel Dynamics in Emerging Institutional Fields. Academy of Management Journal, 52, 355–380, https://doi-org.manchester.idm.oclc.org/10.5465/amj.2009.37308255

Purdy, J., Ansari, S., and Gray, B. (2019). ‘Are Logics Enough? Framing as an Alternative Tool for Understanding Institutional Meaning Making.’, Journal of Management Inquiry, 28(4), pp.409-419. https://doi-org.manchester.idm.oclc.org/10.1177/1056492617724233 (Original work published 2019)

Quattrone, P. (2015). ‘Governing Social Orders, Unfolding Rationality, and Jesuit Accounting Practices: A Procedural Approach to Institutional Logics: A Procedural Approach to Institutional Logics.’ Administrative Science Quarterly, 60(3), pp.411-445. https://doi-org.manchester.idm.oclc.org/10.1177/0001839215592174

Rao H, Monin P, Durand R (2003) Institutional change in Toque Ville: Nouvelle cuisine as an identity movement in French gastronomy. American Journal of Sociology 108(4): 795–843.

Rappaport A (1986), Creating Shareholder Value: The New Standard for Business Performance, New York The Free Press

Reay, T., & Hinings, C. R. (2009). Managing the Rivalry of Competing Institutional Logics. Organization Studies, 30(6), 629-652. https://doi-org.manchester.idm.oclc.org/10.1177/0170840609104803 (Original work published 2009)

Schneiberg, M. & Soule, S.A. (2005).Institutionalization as a contested, multilevel process: The case of rate regulation in American fire insurance. In G.F. Davis, D. McAdam, W.R. Scott & M.N. Zald (eds.), Social movements and organization theory (pp. 122–160). Cambridge, U.K.: Cambridge University Press.

Scott, W (1995), Institutions and Organizations. Ideas, Interests and Identities. Sage. ISBN: 978-142242224

Scott, W. Richard (1994) 'Conceptualizing organizational fields: Linking organizations and societal systems' in System rationalitatund Partial interesse. H. Derlien,U. Gerhardt, and F. Scharpf (eds),203–221. Baden-Baden: Nomos Verlagsgescellschaft.

Scott, W. Richard. (1991). 'Unpacking Institutional Arguments.' Pp. 164–182 in The New Institutionalism in Organizational Analysis, edited by Paul J. DiMaggio and Walter W. Powell. Chicago: The University of Chicago Press.

Scott, W. Richard. (1995: 2001). Institutions and Organizations. Thousand Oaks, CA: Sage Publications.

Scott, W.R. (2008), Institutions and Organizations: Ideas and Interests, 3rd ed. Sage Publications, Los Angeles, CA

Selznick, P. (1949). TVA and the grass roots: A study in the sociology of formal organization. Berkeley, CA: University of California Press.

Seo, M.-G., and W. E. D. Creed (2002) “Institutional contradictions, praxis, and institutional change: A dialectical perspective.” Academy of Management Review, 27: 222-247.

Sine WD, David RJ (2003) Environmental jolts, institutional change, and the creation of entrepreneurial opportunity in the US electric power industry. Research Policy 32(2): 185–207.

Sisson, A and Rogers, D. (2020) International Encyclopedia of Human Geography (Second Edition), Elsevier

Stockhammer E (2004)., ‘Financialisation and the Slowdown of Accumulation’, Cambridge Journal of Economics, 28, pg. 719-741

Strang, D., & Soûle, S. A. (1998). Diffusion in organizations and social movements: From hybrid corn to poison pills. In J.Hagan & K. Cook (Eds.), Annual review of sociology, vol. 24: 265-290. Palo Alto, CA: Annual Reviews.

Sullivan, N and O’Dwyer, B (2015) ‘The structuration of issue-based fields: Social accountability, social movements and the Equator Principles issue-based field’, Accounting, Organizations and Society, 43,pp. 33-55, ISSN 0361-3682, https://doi.org/10.1016/j.aos.2015.03.008.

Swidler, A. (1986). ‘Culture in action: Symbols and strategies.’ American Sociological Review, 51, pp.273–286.

Thornton P. H., Ocasio W., Lounsbury M. (2012). *The institutional logics perspective: A new approach to culture, structure and process*. New York, NY: Oxford University Press.

Thornton, P. H., & Ocasio, W. (2008). Institutional logics. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), The Sage handbook of organizational institutionalism: 99-129. Thousand Oaks, CA: Sage.

Thornton, P. H., and W. Ocasio (1998: 1999) “Institutional logics and the historical contingency of power in organizations: Executive succession in the higher education publishing industry, 1958–1990.” American Journal of Sociology, 105: 801-843.

Thornton, Patricia & Jones, Candace & Kury, Kenneth. (2005). Institutional Logics and Institutional Change in Organizations: Transformation in Accounting, Architecture, and Publishing. Research in the Sociology of Organizations. 23. 10.1016/S0733-558X(05)23004-5.

Tripsas, M., G. Gavetti. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. Strategic Management J. 21(10/11) 1147–1161.

UN (2017) Special rapporteur on adequate housing as a component of the right to an adequate standard of living, and on the right to non-discrimination in this context. Available at: https://www.ohchr.org/en/issues/housing/pages/housingindex.aspx (accessed 27th November 2024).

Van der Zwan, N. (2014) Making sense of financialization, Socio-Economic Review, 12 (1), pp. 99–129, https://doi.org/10.1093/ser/mwt020

Wang, Y. 2015. "The rise of the 'shareholding state': financialization of economic management in China." Socio-Economic Review 13(3), 603–625.

White, H. 1992. Identity and control: A structural theory of social interaction. Princeton, NJ: Princeton University Press.

Wójcik, Dariusz, and Camilleri, James. 2015. "'Capitalist tools in socialist hands'? China Mobile in global financial networks." Transactions of the Institute of British Geographers 40(4), 464–478

Wooten, M and Hoffman, A. (2008) Organizational fields: Past, present and future R. Greenwood, C. Oliver, K. Sahlin, R. Suddaby (Eds.), The SAGE handbook of organizational institutionalism, Sage Publications, London (2008), pp. 130-147

Wooten, M. and Hoffman, A.J., (2017). Organizational fields: Past, present and future. The Sage handbook of organizational institutionalism, pp.55-74.

Wu, F, Chen, J, Pan, F, Gallent, N and Zhang, F. (2020). “Assetization: the Chinese path to housing financialization.” Annals of the American Association of Geographers. https://doi.org/10.1080/24694452.2020.1715195.

Wu, Fulong. (2002). “China's changing urban governance in the transition towards a more market-oriented economy.” Urban Studies 39(7), 1071–93.

Wu, Fulong. (2018). “Planning centrality, market instruments: governing Chinese urban transformation under state entrepreneurialism.” Urban Studies 55(7), 1383–99.

Zhao, H. (2009) ‘A comparative study of the Chinese model and several major development models in the world today’, Red Flag Manuscript, 2009 (22)

Zietsma, C and Groenewegen, P and Logue, D. and Hinings, C. (2017) ‘Field or Fields? Building the Scaffolding for Cumulation of Research on Institutional Fields.’, Academy of Management, 11, pp. 391–450, https://doi.org/10.5465/annals.2014.0052

Zott, C. and Amit, R. (2010), ‘Business model Design: An activity system perspective’, Long Range Planning, 43 (2–3), pp. 216-226