Catalysts for climate solutions: Corporate responses to venture capital financing of climate-tech startups*

Shirley Lu[†], George Serafeim[‡], and Simon Xu[§]

This draft: April 25, 2025
First draft: November 2, 2023

Abstract

This study investigates whether signals from private capital markets in the form of VC financing of climate-tech startups generate information spillovers that influence public incumbent firms' focus on climate solutions. Applying large language models to the 10-K Business Description section to measure this focus, we find that incumbents in similar product markets as VC-backed startups increase their engagement in climate solutions products and services—particularly when the investment reflects strong commercial prospects or involves financially motivated investors. The response is more pronounced among incumbents with a pre-existing focus on climate solutions, and stock prices react positively when these firms also share an industry with the startup. We validate our findings using an instrumental variables approach that addresses endogeneity by leveraging variation in state-level capital gains taxes. Overall, the results highlight VC financing as an informative signal of market demand in uncertain environments, catalyzing incumbent firms' engagement with climate solutions.

JEL Classification: G24; G31; Q54; Q55
Keywords: climate change opportunities, climate solutions, venture capital, startups, climate-tech, generative AI, machine learning

[*]We greatly appreciate the comments of Brian Baik, Andrea Fosfuri, Jonas Heese, Jessica Jeffers, Sofia Johan, Josh Lerner, Ramana Nanda, Edward Riedl, Mike Toffel, Peter Tufano, and seminar participants at Harvard Business School, NUS Business School, Columbia Junior Accounting Conference 2024, Shanghai-Edinburgh-London-Cape Town Green Finance and Accounting Conference 2024, GRASFI 2024, CICF 2024, Oxford Sustainable Private Markets Conference 2024, Yale Initiative on Sustainable Finance Annual Symposium 2024, Boca-ECGI Corporate Finance and Governance Conference 2024, Owners as Strategists Conference 2024, TJAR Conference 2024. We thank Anna Bialas and MarcAntonio Awada for data assistance. All errors are our own.

[†]Harvard Business School, Harvard University, Boston, MA 02163, United States; email slu@hbs.edu.
[‡]Harvard Business School, Harvard University, Boston, MA 02163, United States; email gserafeim@hbs.edu.
[§]Harvard Business School, Harvard University, Boston, MA 02163, United States; email sxu@hbs.edu.

1. Introduction

As the global economy moves toward reducing carbon emissions, climate solutions—products and services that develop or deploy technologies in a transition to a low-carbon economy—have gained increasing attention. Achieving widespread adoption of these solutions by organizations and their customers requires substantial investment, technological innovation, and successful commercialization of climate-related technologies ("climate-tech") (Henderson & Serafeim, 2020). Climate-tech startups play a crucial role in the shift towards a low-carbon economy due to their development of technological innovations aimed at reducing carbon emissions (Noailly & Smeets, 2015). While startups typically pioneer new technologies (Acs & Audretsch, 1987; Schoonhoven, Eisenhardt, & Lyman, 1990; Stuart, Hoang, & Hybels, 1999; Tushman & Anderson, 1986), incumbent firms, with their financial and managerial resources, are often better equipped to commercialize these technologies (Cardinal & Opler, 1995; Gans, Hsu, & Stern, 2002; Glaeser & Lang, 2024; Teece, 1986). However, the market, regulatory, and technological uncertainties surrounding climate solutions imply that investments in this sector may be perceived as high-risk, potentially leading incumbent firms to be reluctant to engage in climate solutions (Blyth et al., 2007; Noailly, Nowzohour, & van den Heuvel, 2022).

In this paper, we hypothesize that venture capital (VC) investments in climate-tech startups serve as an informative signal ("VC signal") to validate the commercial potential of climate solutions.[1] In the face of high uncertainties, if managers from incumbent firms learn from the VC signal about the demand and commercial potential of climate solutions, we expect them to increase their focus on climate solutions to seize the business opportunity.[2] For example, the commercial potential of plant-based meat alternatives—which offer lower emissions relative to traditional meat—is subject to significant uncertainty due to factors such as consumer taste preferences, regulatory pressure to reduce emissions, and the high costs associated with emerging production technologies. In 2011, the startup Beyond Meat received Series A

[1]Prior research finds that VC and private equity (PE) investments generate information spillovers that positively affect local economic growth (Samila & Sorenson, 2011) and industry performance (Bernstein, Lerner, Sorensen, & Strömberg, 2017). More directly, Aldatmaz and Brown (2020) show that PE activity increases productivity, employment, and capital expenditures among incumbent public firms in the same industry. Anecdotal evidence from a recent McKinsey report further supports the idea that incumbents look to emerging climate-tech trends as signals of new business opportunities (Granskog, Patel, Gupta, & Helmcke, 2025)—opportunities that VC investments can often help signal.

[2]Prior literature documents that managers can learn from the market, especially on information that needs to be aggregated from different sources, much like information about product market demands (Bond, Edmans, & Goldstein, 2012; Chen, Goldstein, & Jiang, 2007; Goldstein & Yang, 2019). Relatedly, Hand (2005) and Armstrong, Davila, and Foster (2006) show that the financial statement information used to value VC-backed private firms is also relevant for public equity market valuations, suggesting that public firms may have incentives to learn from signals originating in the private capital market.

funding, with one of the VC investors, Kleiner Perkins, describing that it "could potentially be a black swan to the entire food and agriculture industry," and that they believe "hundreds of entrepreneurs will follow".[3] In the years that followed, we observe both an increase in Google search trends for plant-based protein and the entry of incumbents such as Pinnacle Foods and Boulder Brands into the plant-based protein market.

Specific to climate solutions, VCs are more likely to have information on market demand, given that most climate technologies are in the startup phase with VCs providing the bulk of the funding (IEA, 2024).[4] Additionally, VC investors not only accelerate the quantity and quality of innovation in startups (Bernstein, Giroud, & Townsend, 2016; Hellmann & Puri, 2000; Howell, Lerner, Nanda, & Townsend, 2020; Kortum & Lerner, 2000) but also provide a certification effect through their monitoring, screening, and due diligence of startups (Gompers, Gornall, Kaplan, & Strebulaev, 2020; Hellmann & Puri, 2002; Hsu, 2004; Kaplan & Strömberg, 2001, 2004; Lerner, 1995; Tian, Udell, & Yu, 2016). In particular, Lerner and Nanda (2020, p. 245) assert that "VCs are naturally drawn to investment opportunities where the ideas can be commercialized". From this perspective, VC financing for climate-tech startups represents informed agents committing capital to support climate technologies, thereby serving as an informative signal for the commercial potential for these innovations.

On the other hand, VC investment may not be a credible signal if incumbent firms are better able to assess the market demand for climate solutions. Additionally, incumbent firms may respond to the threat of a startup's climate solutions via other means. Cunningham, Ederer, and Ma (2021) investigate how incumbent firms react to potential new market entrants through "killer acquisitions", specifically to halt the target's innovation projects and proactively thwart future competition. Incumbent firms may also lack the incentive and capability to transition to climate solutions given challenges with organizational inertia (Henderson, 1993). Therefore, it remains an empirical question whether incumbent firms will increase their focus on climate solutions in response to VC financing of climate-tech startups.

To answer this question, we study the impact of VC financing of climate-tech startups on "similar" incumbent firms' focus on climate solution products or services. To measure firms' focus on climate solutions, we use the *CS measure* developed by Lu, Serafeim, Xu, and Awada (2024), which is based on data derived from large language models (LLM) applied on the "Business Description" section of Item 1 in U.S. publicly listed firms' 10-K filings. This

[3] `https://www.kleinerperkins.com/case-study/beyond-meat/`.

[4] In 2021, climate-tech startups raised $53.7 billion from VC and PE (BloombergNEF, 2021). The growth rate of capital invested in climate-tech startups between 2015 and 2021 exceeded 150% (Cornelli, Frost, Gambacorta, & Merrouche, 2023).

section is particularly suitable for our analysis as it provides a legally mandated, detailed account of companies' products and services, reducing the likelihood of misinformation (i.e., greenwashing) and offering a standardized text for LLM analysis. To identify firms' climate solution products and services, Lu, Serafeim, et al. (2024) fine-tune a Generative Pre-trained Transformer (GPT) model using a labeled dataset to identify sentences related to climate solutions in 13 (out of 25) GICS industry groups where climate technologies and practices codified by the Project Drawdown can be found. This process enables the construction of the *CS measure*, defined as the ratio of climate solutions sentences to the total number of sentences in Item 1, which serves as a proxy for a firm's focus on climate solutions in its product and service offerings.[5]

To investigate how VC financing of climate-tech startups affects incumbent firms' *CS measure*, we first match all incumbent firms that are "similar" to a given startup, as these are the incumbents most likely to respond. Given that startups can operate across industry boundaries and their products may resemble those of incumbents from various industries, we adapt the methodology of Hoberg and Phillips (2016) to assess the degree of product market overlap between incumbents and startups.[6] Using data from PitchBook, we identify a set of startups that fall under the category of climate-tech. Then, for each startup, we compute a set of similarity scores by conducting textual analysis on the business descriptions of the startup and the business descriptions of U.S. listed companies, using the 10-K filings submitted at the fiscal year-end preceding the deal date of the VC financing round. We define the set of *similar* incumbent firms as those with similarity scores in the top one percentile.

We employ a stacked difference-in-differences (DiD) regression to evaluate the impact of VC financing rounds on the *CS measure* of incumbent firms (Baker, Larcker, & Wang, 2022; Cengiz, Dube, Lindner, & Zipperer, 2019; Gormley & Matsa, 2011). We focus on a four-year window around VC financing rounds of climate-tech startups and create clean "cohorts" of VC financing rounds by comparing the changes in the *CS measure* of similar incumbents ("treated") to a set of never-treated non-similar incumbents ("control") within the same 6-digit GICS industry code. Conceptually, these control firms should have a similar information environment regarding the general demand for climate solutions (e.g., regulatory shocks related to decarbonization), except that treated firms are more likely to respond to the VC signal

[5] We report results also using an alternative measure of climate opportunities derived from a less advanced natural language processing model from earnings calls (Sautner, van Lent, Vilkov, & Zhang, 2023) and find directionally similar results.

[6] Hoberg and Phillips (2016) examine product market similarity among listed firms, whereas we adapt their approach to measure product similarity between a startup and a listed incumbent firm.

given more similar product markets with the startup. Our DiD sample covers the 2005-2021 period and includes 3,961 VC financing rounds involving 1,932 unique startups and 1,676 unique treated and control incumbents, resulting in 59,898 incumbent firm-year observations. Besides conditioning on firm × cohort and year × cohort fixed effects, the analysis controls for an array of time-varying variables that may influence incumbent firms' focus on climate solutions.

We find that similar incumbent firms increase their focus on climate solutions more than non-similar firms in response to VC financing of climate-tech startups. This finding is consistent with the interpretation that VC financing serves as a signal for the commercial potential of climate technologies, which induces similar incumbent firms to boost their focus on providing climate solution products and services. Estimates imply an economically significant effect. For example, consider two otherwise identical firms, except that one is identified as a similar incumbent firm in a VC financing round, while the other is not. The coefficient estimates indicate that in the years following the VC financing round, the similar firm increases its *CS measure* by 0.27 percentage points more than the non-similar firm. This increase corresponds to approximately 37% of the sample mean of *CS measure* and 11% of its sample standard deviation. Furthermore, estimating a dynamic model reveals no differential trends in the *CS measure* between similar and non-similar firms prior to VC financing rounds, with the observed increase in the *CS measure* for similar firms occurring only after the VC financing.

Next, to increase our confidence that incumbent firms respond because VC investment signals the commercial potential for climate solutions, we focus on cross-sectional variation where this signal is likely to be stronger. First, we expect the signal to be stronger if the VC investment demonstrates a higher commercial potential for climate solutions. Consistent with this expectation, the increase in *CS measure* is significantly larger in deals with a higher deal size and post-investment valuation, and when a startup is already generating revenue—factors that suggest investors see a significant commercial potential for the startup (Burt, Harford, Stanfield, & Zein, 2023; Pham, Rezaei, & Zein, 2023). Moreover, similar incumbent firms exhibit a more pronounced increase in their *CS measure* if the startup is funded by VCs with a track record of superior investment performance and a commitment to investing in the climate-tech sector. When we separate the investor syndicate into traditional VCs and impact-focused investors, the main effect only holds for traditional VCs that are financially motivated. In contrast, the effect is attenuated for impact-focused investors, consistent with a weaker signal for commercial potential when the investor pursues non-financial objectives.

Second, we expect a stronger response from similar incumbent firms when the VC investment is more visible, and hence is more likely to gain attention from incumbent firms. We find that the increase in *CS measure* of similar incumbent firms is significantly larger in startup deals with more new investors and for startups with higher media coverage leading up to the deal. The finding that the strength of the association varies predictably with the strength of the signal further provides some comfort that the signaling mechanism is likely to be an important driving force for the documented association.

We conduct three robustness tests to directly address potential endogeneity concerns related to reverse causality, where VC investments occur after observing incumbents' focus on climate solutions, and confounding shocks, such as market or technological developments, that might simultaneously drive both VC financing of climate-tech startups and increases in similar incumbent firms' focus on climate solutions.[7] First, we implement a two-stage least squares (2SLS) approach using variation in state-level capital gains taxes faced by VC firms as an instrument for the likelihood that startups receive VC financing.[8] This variation is plausibly exogenous, as prior research shows that changes in capital gains tax rates significantly affect VC firms' investment behavior (Dimitrova & Eswar, 2023; Keuschnigg & Nielsen, 2003, 2004; Lerner & Nanda, 2020), influencing the probability that startups receive VC financing without directly affecting incumbent firms' climate strategies. Instrumenting our treatment variable with state-level capital gains taxes, we find that the effect of VC financing of climate-tech startups on similar incumbent firms' climate solution focus remains positive and statistically significant. These findings provide evidence that the observed incumbent response reflects an exogenous reaction to the VC signal, rather than being driven by endogenous selection or correlated omitted variables.

Second, to provide further support that our results are not driven by confounding shocks, we conduct a set of complementary tests. We begin by applying the robust inference framework of Rambachan and Roth (2023), which bounds the relative magnitude of post-treatment deviations from parallel trends compared to pre-treatment violations due to confounding shocks. Using this approach, we confirm that our results remain statistically significant under

[7]The existence of other information alone does not violate the signaling story. Rather, the signaling interpretation suggests that VCs, given their relevant expertise, provide a more credible signal of the commercial potential for climate solutions relative to other information. The concern about confounding arises if we observe increases in incumbents' climate solutions even in the absence of VC investment, suggesting that other confounding shocks may be driving the results.

[8]This instrumental variable affects the availability of the VC signal, but does not imply that any increase in VC investment mechanically causes incumbent responses. For example, when capital gains tax rates are sufficiently high such that VC investment is deterred entirely, the signal becomes unavailable, limiting the ability of incumbents to learn from it.

standard benchmark deviations from the parallel trends assumption. We then examine whether the effects of VC financing vary across periods of high versus low unexpected climate change concerns (Ardia, Bluteau, Boudt, & Inghelbrecht, 2023), and find that the relationship between VC financing and incumbent firms' focus on climate solutions persists across both periods. Additionally, we split the sample based on energy market conditions—such as oil, natural gas, and solar prices—and again find no significant differences in the estimated effects. The fact that the strength of the association does not vary according to these variables but it varies systematically with the strength of the signal from the VC investment further increases our confidence that incumbents respond to the VC signal rather than a confounding effect driving the association.

Third, we rule out mergers and acquisitions (M&A) as a driver of the observed increase for similar incumbent firms' *CS measure*. Event study analysis reveals no significant changes in M&A activity for similar incumbent firms before or after VC financing rounds, including acquisitions of climate-tech startups. These results indicate that the increase in incumbent firms' focus on climate solutions is neither driven by efforts to position themselves as potential exit options for VC-backed startups nor by "killer acquisitions" intended to preempt future competition (Cunningham et al., 2021).

In additional analysis, we examine which incumbent firms exhibit a more pronounced response to these VC signals. Specifically, incumbents operating in the same industry as the startup exhibit a stronger response, likely due to overlapping customer bases that make the VC signal more directly relevant to their markets. Additionally, incumbents with a pre-existing focus on climate solutions also show a stronger response, as these firms possess complementary assets that enable them to capitalize on the increased market demand for climate solutions. While it is possible that competitive threats from startups could also drive incumbents to increase their focus on climate solutions, our evidence suggests that the VC signal channel dominates. Analyzing stock price reactions around VC financing dates, we find that similar incumbents operating in the same industry as the startup experience significantly higher cumulative abnormal returns (CARs) if they already have a pre-existing focus on climate solutions compared to those without such a focus. This positive market reaction is more consistent with the VC investment being perceived as a signal of market potential rather than as a heightened competitive threat.

We also examine the investment implications of incumbent firms' increased focus on climate solutions. One potential concern is that the language-based measure of climate solutions might be disconnected from actual investments in which case incumbent firms are talking about climate solutions but not allocating resources to them. Lu, Serafeim, et al. (2024) validate the *CS measure* by showing that it correlates with measures of green patents and green revenues, where a one standard deviation increase in *CS measure* is associated with a 0.5 standard deviation increase in green revenue and green patent percent. However, we add our own evidence to this question by investigating the investment implications of the VC signal and validating that the *CS measure* captures an increased real product focus on climate solutions. Using VC investments in climate-tech startups as an instrumental variable, we examine how the VC-induced increase in *CS measure* affects firm investments. Our findings reveal a positive and statistically significant relationship; that is, VC-induced increases in *CS measure* lead to higher firm investments measured across three key financial metrics: capital expenditures, research and development expenses, and a decrease in dividend payouts (a proxy for higher reinvestment). This analysis demonstrates that, in response to the VC signal, similar incumbent firms focus more on climate solutions, evidenced by an increase in investments and research and development.

Finally, our results remain robust when using different methodological specifications. First, to validate our measure of incumbents' climate solutions, we repeat our analysis using firm-level climate change opportunities measure developed by Sautner et al. (2023), which are based on earnings call attention rather than 10-K filings. Our findings reveal that similar incumbent firms increase their exposure to climate-related opportunities and show more positive sentiment towards these opportunities in response to VC financing rounds, while their exposure to regulatory and physical risks, as well as uncertainty or negative sentiment towards these opportunities, does not significantly change. Second, the results hold when we use alternative stacked DiD specifications such as employing different definitions for the control group, and when we use propensity score matching to address the potential nonrandom assignment of incumbent firms into treated and control groups. Third, the results remain robust when using a staggered two-way fixed effects DiD and implementing the procedure by de Chaisemartin and D'Haultfoeuille (2020) to address potential biases arising from heterogeneous treatments in DiD settings.

Our study contributes to the literature on how firms respond to economic signals that inform potential market demand in uncertain environments. A large body of research shows that when uncertainty or information frictions are high, market participants rely on external signals to guide decision-making. For example, peer disclosures lower the cost of capital for first-time private issuers (Shroff, Verdi, & Yost, 2017), and public firm disclosures facilitate trading in the patent market by revealing value-relevant information in a highly opaque environment (Kim & Valentine, 2023). In other settings shaped by uncertainty, acquirers strategically disclose information during M&A negotiations to shape perceptions that lead to lower acquisition costs (Kim, Verdi, & Yost, 2020), new entrants reduce uncertainty about optimal capital structure in concentrated markets by mimicking those of incumbents (Bernard, Kaya, & Wertz, 2021), and entrepreneurs use information from local IPOs to assess the viability of launching new ventures (Barrios, Choi, Hochberg, Kim, & Liu, 2023). We add to this literature by showing that VC financing of climate-tech startups serves as a credible signal of commercial potential, prompting similar public firms to increase their product focus on climate solutions. In doing so, we respond to calls for research on how managers use external signals to allocate capital across competing investment opportunities under uncertainty (Ferracuti & Stubben, 2019).

We further contribute to the accounting literature on VC activity in private capital markets by examining how VC investments in startups generate information spillovers that shape the strategic decisions of public firms. Prior research has shown that private firm disclosures increase the likelihood of receiving PE and VC financing (Baik, Berfeld, & Verdi, 2025), and that PE- and VC-backed firms, at the time of their IPO, exhibit greater transparency (Katz, 2009), are less likely to manage earnings (Morsfield & Tan, 2006), and are more likely to issue earnings guidance (Allee, Christensen, Graden, & Merkley, 2021). We extend this literature by showing that VC financing of startups influences non-portfolio public firms through product-market information spillovers, which in turn shape investment-relevant outcomes. This perspective also connects our study to the broader literature on information spillovers between private and public firms. For example, greater accounting comparability facilitates valuation spillovers from public to private firms (Bourveau, Chen, Elfers, & Pierk, 2023), private firm disclosure transparency affects capital flows to public firms (Kim & Olbert, 2022), and private firms are more responsive to investment opportunities in industries with greater public firm presence (Badertscher, Shroff, & White, 2013). Our findings add to this literature by documenting that capital allocation decisions in private markets, specifically VC financing, can act as informative signals that influence public firms' business decisions.

Finally, our study contributes to the growing literature on the forces that shape corporate climate transitions. While much of the existing research adopts a risk-based perspective, focusing on firms' greenhouse gas emissions or exposure to transition risks (Bolton & Kacperczyk, 2021; Sautner et al., 2023), we build on emerging work that examines how firms pursue opportunities arising from climate mitigation (Cohen, Gurun, & Nguyen, 2020; Lu, Riedl, Serafeim, & Xu, 2024). In addition, prior studies emphasize the role of regulation in driving climate-related corporate action, such as disclosure mandates (Downar, Ernstberger, Reichelstein, Schwenen, & Zaklan, 2021; Tomar, 2023) and carbon pricing policies (Martin, Muûls, & Wagner, 2016; Teixidó, Verde, & Nicolli, 2019). We extend this literature by showing that VC investment acts as a market-based signal of commercial potential, increasing incumbent firms' engagement in climate solutions. Our findings suggest that, beyond regulatory pressure or risk mitigation, market-based signals about future demand play a role in shaping corporate climate action.

2. Data, sample, and variables

2.1. Startup sample

Our sample consists of startups headquartered in the U.S. identified with data from PitchBook. This dataset has comprehensive coverage of various aspects of startup financing rounds, including details such as timing, stage (e.g., Seed, Series A, B, C, etc.), investment amount, and the identity of investors involved in each round. PitchBook further categorizes startups into "verticals" based on their technological orientation (e.g., FinTech, Nanotechnology, Software-as-a-Service, etc.). These verticals group startups into clusters that concentrate on a shared niche or specialized market.[9] Our analysis specifically focuses on startups falling under the "Climate Tech" or "CleanTech" verticals.[10] We consider VC financing rounds taking place between 2005 to 2021. To be included in our sample, a financing round must meet the following criteria: 1) it is explicitly identified in the PitchBook database as a "Venture Capital" round with at least one investor in the syndicate identified as a VC investor by PitchBook;[11] 2) it must have non-missing data for deal size and deal date; and 3) it must involve the raising of new equity (debt-only and secondary-sale rounds are excluded).

[9]A single vertical may be comprised of companies that span multiple industries. PitchBook explains the differences between verticals and industry classifications here: `https://pitchbook.com/what-are-industry-verticals`.

[10]Based on PitchBook's definition, the "Climate Tech" vertical includes "companies developing technologies intended to help mitigate or adapt to the effects of climate change. The majority of companies in this vertical are focused on mitigating rising emissions through decarbonization technologies and processes. Applications within this vertical include renewable energy generation, long duration energy storage, the electrification of transportation, agricultural innovations, industrial process improvements, and mining technologies, among others." Similarly, the "CleanTech" vertical includes "developers of technology which seeks to reduce the environmental impact of human activities or to significantly reduce the amount of natural resources consumed through such activities."

[11]This restriction excludes VC rounds financed purely by individuals, angel groups, accelerators/incubators, crowdfunding investors, etc.

Panel A of Table 1 presents summary statistics for climate-tech startup deals sorted by year. Our sample consists of 3,961 deals involving 1,932 unique startups. The observed trends in the number of deals and the average deal size per year align with patterns documented in previous studies (Cornelli et al., 2023; Gaddy, Sivaram, Jones, & Wayman, 2017; van den Heuvel & Popp, 2023). Specifically, the period from 2005 to 2011 witnessed an initial boom in climate-tech investments with a surge in both the number of deals and average deal size. Subsequently, there was a contraction in deal activity from 2012 to 2014. Following this downtrend, the second boom period occurred from 2015 to 2021, characterized by substantial investment inflows, with the average deal size peaking at $41 million in 2021. The average number of investors per deal also exhibits an upward trend beginning from the second boom period.

2.2. Identifying similar incumbent firms

Our objective is to identify, from the entire set of listed incumbent firms, those that are most similar to a particular startup. We follow the methodology of Hoberg and Phillips (2016), which suggests assessing the degree of overlap between the features of the products offered by the incumbents and those of the startup. While Hoberg and Phillips (2016) concentrate on product similarity among listed firms, we adapt their approach to measure the product similarity between a startup and a listed incumbent firm.

PitchBook provides a business description of each startup that summarizes the startup's primary products and target customer base.[12] Although these descriptions generally do not exceed a paragraph, PitchBook structures them in a way that comprises highly informative words.[13] We parse each description to extract a set of keywords that closely represent a startup's product characteristics. Following Hoberg and Phillips (2016), the keywords consists of rare and proper nouns identified in the descriptions. Rare nouns are those appearing in less than 25% of descriptions, while proper nouns are capitalized more than 90% of the time. Geographic location names, including countries, U.S. states, and largest cities, are excluded.

The set of extracted keywords is used in conjunction with the business description of each incumbent firm, found in Item 1 of its 10-K filings, to assess the extent to which the

[12] PitchBook primarily acquires information from the startup's website, and their research team manually crafts a description based on that data. In cases where a website is unavailable, PitchBook relies on information from press releases and investor portfolios to provide a summary of the startup's operations.

[13] The standard structure of a business description typically commences with a succinct overview of the startup's core product or service. It then delves into more detailed aspects, describing features and benefits while emphasizing the value that the product or service offers to its target customer base. If information is available, PitchBook also includes a description of the startup's business model and a statement of purpose that articulates the reason for the company's establishment.

incumbent firm's existing product features share similar characteristics with those of the startup. Following Hoberg and Phillips (2016), we exclude firms without valid Compustat data, firms with nonpositive sales, and firms with assets of less than $1 million. Since the product information of a listed firm changes over time due to annual updates in its 10-K filings, we rely on the most recent 10-K filed at the fiscal year-end preceding the deal date of the VC financing round.[14] This approach, by identifying similar firms before the startup secures financing, serves to mitigate look-ahead bias as it ensures that we do not inadvertently capture the product characteristics of firms that, ex post, choose to become similar to the startup.

We compute a similarity score for each pair of a startup (j) and an incumbent firm (i), reflecting the proportion of keywords shared between the startup's description and the incumbent's Item 1 Business Description. Following Loughran and McDonald (2011), we assign a weight to each keyword k, denoted as w_k, defined as the log inverse of its frequency across all startup descriptions in our sample:[15]

$$w_k = \log(N/f_k), \tag{1}$$

where N is the total number of startups across all VC financing rounds and f_k is the number of startups whose descriptions include the keyword k. Formally, the similarity score between startup j and incumbent i is given by:

$$\textit{Similarity score}_{j,i} = \frac{W \cdot (S_j \circ S_i)}{W \cdot S_j}, \tag{2}$$

where W is the vector containing the weights w_k, S_j is the vector containing the keywords associated with startup j, and S_i is the vector containing the keywords extracted from all startup descriptions that are found in incumbent firm i's Item 1 Business Description. $S_j \circ S_i$ represents a vector that contains the shared set of keywords between startup j and incumbent firm i.[16] The similarity score ranges from 0 to 1 and signifies the extent to which the incumbent

[14]For instance, if a firm has a fiscal year ending in September and a startup secures VC financing in August 2015, we would utilize the 10-K filed as of the fiscal year-end in September 2014.

[15]This weighting scheme means that less frequent words convey more information about the startup's business.

[16]The denominator of the similarity score in Equation (2) includes only S_j and not S_i, differing from Hoberg and Phillips (2016). If we were to also incorporate S_i in the denominator, it would imply that incumbent firms with diverse product features (i.e., those with longer Item 1 Business Descriptions) are more dissimilar to a given startup compared to more focused incumbent firms (i.e., those with shorter Item 1 Business Descriptions) that share the same number of keywords with the startup. By dividing only by S_j, we ensure equal treatment of these two incumbent firms in terms of similarity to the startup.

firm's product market overlaps with that of the startup, with a higher score indicating greater similarity.

We define the set of "similar" incumbent firms for a specific startup as the listed incumbent firms ranking within the top one percentile of similarity scores with respect to the given startup j. In the following section, we examine the characteristics of this set of similar incumbent firms to rationalize their selection.

2.2.1. Characteristics of similar incumbent firms

Figure 1 plots the average similarity score across all startups for a given rank.[17] The vertical dashed line represents the average cutoff rank for the top one percentile of similarity scores. As depicted in Panel A, the set of similar incumbent firms corresponding to the top one percentile of similarity scores comprises approximately 32 incumbent firms. The average similarity score is as high as 0.75 for the first-ranked (most similar) incumbent firm and drops to 0.57 for the 32nd ranked incumbent firm. Panel B shows that moving down from the first to the 32nd rank leads to a significant change in the average similarity score. However, from the 32nd rank onwards, the graph remains roughly flat, with the average change in similarity averaging only 0.09%. Thus, the top one percentile of incumbent firms constitutes a group of similar firms where the addition of further firms to the set has negligible effects on the average similarity score.

As further validation, we assess the industry overlap between the GICS industry group of the set of similar incumbent firms and startups. Although startups are typically not assigned a GICS industry code, PitchBook assigns its own industry code to each startup.[18] Panel B of Table 1 displays the distribution of startups across industry sectors based on the PitchBook industry classification. The majority of startups are in the Consumer Products and Services sector, followed by Energy, and Information Technology. We manually match each startup's PitchBook industry code to at least one 4-digit GICS industry group based on the descriptions provided by the respective industry taxonomy.[19] On average, we observe that 34% of similar incumbent firms share the same 4-digit GICS industry group as the startup. Thus, our

[17]For example, at rank 1, the graph shows the average similarity score across all startups for incumbent firms with the highest similarity score.

[18]PitchBook's industry classification is based on the GICS, making them highly comparable.

[19]To illustrate, consider the PitchBook industry code 1.1.1, denoted as "Aerospace and Defense," with the definition "Manufacturers of equipment, parts or products related to civil or military aerospace and defense. Includes aircraft parts, firearms, and other munitions." This PitchBook industry is matched with the GICS industry labeled "Aerospace & Defense," which has a 4-digit GICS industry group code of 2010. The PitchBook industry definition closely aligns with the GICS industry definition: "Manufacturers of civil or military aerospace and defense equipment, parts or products. Includes defense electronics and space equipment." In general, a PitchBook industry code could be matched to more than one GICS industry group.

methodology is effective at capturing the set of incumbent firms that are most similar to startups, as there is a modest degree of industry overlap between them.

In another validation test, we examine the likelihood that similar incumbent firms belong to the same GICS industry. If our methodology is effective in identifying incumbent firms similar to startups, there should be a reasonable degree of overlap in GICS codes among these similar incumbents. Panel C of Table 1 presents the distribution of similar incumbent firms across GICS industry sectors. The top three industries where similar firms belong to are Industrials, Information Technology, and Energy—coinciding with the top three sectors for most startups. On average, 59% of similar incumbent firms share the same 6-digit GICS industry code. While not directly comparable, our figure is similar to that of Hoberg and Phillips (2016), who document that two listed companies with the same text-based industry classification are 44% likely to belong to the same 3-digit SIC code.

2.3. Climate solutions large language model

To measure incumbent firms' focus on climate solution products and services, we rely on a measure derived from a GPT model fine-tuned to detect climate solutions sentences in the "Business Description" section of 10-K filings from the Securities and Exchange Commission's (SEC) EDGAR database (Lu, Serafeim, et al., 2024). The labeling for climate solution sentences is based on Project Drawdown, which contains a list of technologies that can reduce greenhouse gases in the atmosphere, and are compiled by a network of scientists and researchers. Our sample period spans fiscal years 2005 to 2021 and covers 13 (out of 25) GICS industry groups that are central to climate solutions, where LLM is more accurate in identifying climate solutions.

GPT is well-suited for this task since separating climate solution sentences from other climate sentences requires more advanced context recognition than other methods such as lexicon-based approaches, and the pre-trained GPT model is more capable of understanding contextual sentences. For example, "We produce electric vehicles" is considered a climate solutions sentence, but "We believe we have a responsibility and opportunity to play a role in the global economic transition to net zero emissions" is not. As a more challenging example, the sentence "Primary fleet EV competitors include Smith Electric, Azure Dynamics, Enova, and EnVision Motor Company" is classified as a climate solutions sentence but "Electric vehicle industry growth has accelerated in the past several years" is not. While both sentences include the climate solution electric vehicles (EV), the former implies the focal firm produces EV and has EV competitors, while the latter merely describes an industry trend without sufficient information to suggest the focal firm produces EV. The fine-tuned climate solutions GPT model achieves an accuracy rate of 84.09% and an F1 score of 0.79, indicating a high level of precision and recall in its predictions.[20]

The climate solutions GPT model is applied to all sentences in 10-K Item 1 in the sample. To capture the relative importance of climate solutions for a given firm-year, the *CS measure* is defined as the number of climate solutions sentences divided by the total number of sentences in 10-K item 1. Previous research has shown that this measure correlates with other measures of climate opportunities, such as green patents, green revenues, and several innovation measures required to develop climate solutions (Lu, Serafeim, et al., 2024). In the Internet Appendix, we provide an extract on the LLM methodology section of Lu, Serafeim, et al. (2024), which provides more details on the construction and labeling of the climate solutions GPT model.

2.4. Control variables

We include a number of variables to control for factors that may affect incumbent firms' focus on climate solutions. We obtain financial information from Compustat and stock prices from CRSP. Controls for firm fundamentals include the natural logarithm of one plus the number of years since the firm is first recorded in the CRSP stock database (*Firm age*), the natural logarithm of total assets (*Firm size*), the natural logarithm of the book-to-market ratio (*Book-to-market*), return on assets (*ROA*), book leverage (*Leverage*), current fiscal year sales divided by previous fiscal year sales minus one (*Sales growth*), and cash divided by total assets (*Cash*). Controls for stock characteristics include the cumulative 12-month return of a stock, excluding the immediate past month (*Momentum*), the annual stock return of the firm (*Stock return*), and the standard deviation of monthly stock returns over the past 12 months (*Stock volatility*). Controls for existing industry concentration include a given firm's sales divided by the total sales of all listed firms in the same SIC2 industry (*Market share*) and Hoberg, Phillips, and Prabhala's (2014) product market fluidity measure (*Fluidity*). Table A.1 in Appendix A describes the control variables in detail.

3. Research design

3.1. Baseline DiD specification

Given the observed bias in conventional two-way fixed effects DiD models (Baker et al., 2022), we adopt the approach recommended in the literature by employing a stacked DiD specification

[20] The F1 score is calculated the harmonic mean of precision and recall. Precision is the percentage of predicted positives that are truly positive. Recall is the percentage of true positives that are predicted as positives.

(Cengiz et al., 2019; Gormley & Matsa, 2011). Specifically, we focus on an event window of four years before to four years after startups' VC financing rounds. For each VC financing round of a given startup (which we call a cohort), we construct cohort-specific "clean" datasets. An incumbent firm is treated in a given VC financing round if it is identified as a similar firm in that round and not in any other rounds in the preceding four years. If a treated firm is classified as a similar firm in multiple startup deals that occur in the same event year, we assign it to the deal where it has the highest similarity score.

To qualify as a control, incumbent firms cannot be identified as a similar firm for any other rounds in the entire event window. Additionally, control firms are matched to the same 6-digit GICS industry code as the treated firms. To define the control sample, our goal is to identify incumbents that are "not" similar to the startup, corresponding to incumbent firms below a certain percentile of similarity scores. However, setting a percentile cutoff that is too low (e.g., the bottom one percentile) limits the full variation in non-similar incumbents. Conversely, setting a percentile cutoff that is too high (e.g., the bottom 80 percentile) risks including too many control firms that may not be non-similar enough. For our baseline specification, we designate an incumbent firm as a control in a given VC financing round if it is in the bottom 20th percentile of similarity scores.[21] We choose this percentile cutoff because, on average, for each VC financing round, each treated firm corresponds to about 5 control non-similar firms after matching based on the 6-digit GICS industry code.[22] Appendix B provides an example of the algorithm used to identify similar and non-similar incumbent firms for a given cohort.

These cohort-specific datasets are then pooled together, and a DiD regression is estimated on the stacked dataset, allowing for firm and year fixed effects to vary by cohort. Formally, we estimate the following regression:

$$CS\ measure_{i,t,c} = \beta\, Top\ Similar_{i,c} \times Post_{t,c} + \gamma X_{i,t-1} + \tau_{i,c} + \rho_{t,c} + \varepsilon_{i,t,c}, \tag{3}$$

where i denotes firm, t denotes year, and c denotes cohort. $CS\ measure_{i,t,c}$ is the percentage of sentences identified as climate solutions to the total number of sentences in firm i's 10-K Item 1 Business Description in year t and cohort c. $Top\ Similar_{i,c}$ is a dummy variable equal to one if firm i is treated in cohort c, and zero otherwise. $Post_{t,c}$ is a dummy variable equal to

[21]Incumbent firms that have never been identified as a similar firm will function as clean controls across all cohorts. Firms treated in later cohorts might appear multiple times since they can also serve as controls for earlier cohorts.

[22]As discussed in Sections 6.2 and 6.3, we examine the robustness of the results with alternative cutoffs and specifications for the control sample.

one for the event year and subsequent four years in cohort c, and zero otherwise. $X_{i,t-1}$ is a vector of control variables. The standard errors are clustered at the firm level. The DiD estimate, β, measures the average treatment effect of climate-tech startups' VC financing on similar incumbent firms' focus on climate solutions.

Firm $\times$ cohort fixed effects ($\tau_{i,c}$) and year $\times$ cohort fixed effects ($\rho_{t,c}$) subsume the main effects for $\textit{Top Similar}_{i,c}$ and $\textit{Post}_{t,c}$, respectively. The inclusion of $\tau_{i,c}$ controls for unobservable incumbent firm characteristics that may influence a firm's focus on climate solutions. For example, long-term institutional shareholders that hold a large stake in certain firms may prioritize engagement efforts on sustainability issues that include a focus on climate solutions (Azar, Duro, Kadach, & Ormazabal, 2021; Naaraayanan, Sachdeva, & Sharma, 2021). Similarly, the introduction of $\rho_{t,c}$ accounts for common time-specific shocks that simultaneously affect both incumbent firms' investments in climate solutions and climate-tech startup deals. For example, if a surge in market demand for a specific green technology impacts both incumbents' and startups' investments in climate solutions in a similar manner, it can be controlled for using $\rho_{t,c}$.

4. Main analyses

4.1. Summary statistics

Figure 2 depicts the distribution of average similarity scores among incumbent firms identified in the top one percentile of the similarity score. The distribution closely mirrors a normal distribution, with scores symmetrically distributed around the mean and no significant concentrations in the tails. This observation suggests that our identification of similar firms avoids a bias towards either an excess of very low similarity scores or an abundance of very high similarity scores. The distribution of scores also indicates that some startups have very few similar incumbent firms, while others have many, aligning with the characterization of startup activity as a spectrum that includes both innovators and imitators (Hellmann & Puri, 2000).

Table 2 presents summary statistics for key variables. Out of the 59,898 firm-year observations used in the DiD regression, approximately 16% belong to the treated group. The mean value of the *CS measure* is 0.724, indicating that, on average, 0.72% of the sentences in an incumbent firm's 10-K Item 1 Business Description are identified as related to climate solutions. The standard deviation is 2.5, signifying considerable variation in the extent to which sentences are labeled as climate solutions across incumbent firms.

4.2. Baseline results

Table 3 reports the estimates of Equation (3). Column (1) includes firm × cohort and year × cohort as the only control variables to maintain the largest possible sample size and to alleviate the concern that including additional covariates could confound estimates of β if they are also affected by the treatment (Gormley & Matsa, 2014). We find that the coefficient on *Top Similar* × *Post* is positive and significant, indicating that similar incumbent firms increase focus on climate solutions more than non-similar incumbent firms in response to climate-tech startup deals. Introducing firm and stock characteristics as additional control variables in columns (2) and (3), respectively, the coefficient on *Top Similar* × *Post* continues to load significantly positively. Column (4) is the most stringent specification, incorporating controls for incumbent firms' existing product market competition. The coefficient estimate in this column indicates that similar incumbent firms increase their *CS measure* by 0.27 percentage points more than non-similar firms in response to climate-tech startups' VC financing rounds. In economic terms, this effect translates to 37% (= 0.265/0.724) of the sample mean of *CS measure* and 11% (= 0.265/2.494) of its sample standard deviation. Taken together, these findings support the hypothesis that VC financing for climate-tech startups stimulates similar incumbent firms to increase their focus on climate solutions.

We also conduct a placebo test to further assess the impact of VC financing rounds on similar incumbent firms' *CS measure*. We estimate 1,000 simulations of the regression in column (4) of Table 3. In each simulation, we randomly assign *Top Similar* across incumbent firms rather than using the actual definition of *Top Similar*. We collect each simulation's estimated coefficient on the placebo term *Top Similar* × *Post*. Figure 3 plots the kernel density distribution of these estimated coefficients and the corresponding p-values. As shown, the majority of the simulated βs are concentrated around zero and not statistically significant at the 10% level, while the "true" β is on the very right tail of the distribution. The results from the placebo test suggest that our findings are unlikely to be driven by random variation in the assignment of *Top Similar*.

4.2.1. Dynamic effects

Our identification is based on the parallel trends assumption, that both treated and control firms exhibit similar trends in *CS measure* prior to VC financing rounds. To validate the assumption that the trends in *CS measure* of the treated and control firms would be the same in the absence of startup deals, we estimate a dynamic version of Equation (3), focusing on the four years preceding and following VC financing rounds as follows:

$$CS\ measure_{i,t,c} = \sum_{\substack{\ell=-4 \\ \ell \neq -1}}^{\ell=+4} \lambda_\ell \, Top\ Similar_{i,c} \times \theta^{\ell}_{t,c} + \gamma X_{i,t-1} + \tau_{i,c} + \rho_{t,c} + \varepsilon_{i,t,c}, \tag{4}$$

where $\theta^{\ell}_{t,c}$ is a dummy variable that equals to one for year ℓ relative to the event year in cohort c, and zero otherwise. The dynamic effects, λ_ℓ, provide event-study style regression estimates that reflect the changes in *CS measure* between similar and non-similar firms over time, both before and after VC financing rounds. We define the year prior to the VC financing round as the reference period, denoted by $\ell = -1$.

Figure 4 shows the dynamic effects from estimating Equation (4). There is no indication of any significant differences in *CS measure* prior to startups' VC financing rounds, which lends support to our assumption that there are no differential responses in incumbents' focus on climate solutions before startup deals. However, beginning in the year of the VC financing round, a gap opens up so that similar firms' *CS measure* is about 0.15% higher compared to non-similar firms. The gap grows wider in the following three years before decreasing in the fourth year. This gradual increase in magnitude aligns with the idea that firms often have a stock of innovations or products that they can deploy when market conditions change, while it also takes time to further invest in and develop new climate solutions. Furthermore, none of the 95% confidence intervals of λ_ℓ in the post-VC financing period overlaps with those in the pre-VC financing period. Overall, the parallel trends assumption is likely to be satisfied in our DiD research design.

4.3. Cross-sectional characteristics by signaling strength

To provide evidence on the signaling effect of VC investment in climate-tech startups, we conduct cross-sectional tests to explore how variations in signaling strength across VC financing rounds affect similar incumbent firms' focus on climate solutions. We augment Equation (3) by including triple interaction terms with deal characteristics that are expected to have larger signaling effects on incumbent firms' *CS measure*: (i) the valuation and financial prospect of the startup; (ii) the type of investors involved; and (iii) the visibility of the startup deal.

4.3.1. Startup valuation and financial prospect

Large VC investments and high startup valuations are often interpreted as indicators of market demand and investor confidence in the startup's potential success. Incumbent firms may see this as a signal that there is a viable market for the commercialization of climate solution products and services, and hence increase focus on climate solutions.

We use three variables to measure startups' valuation and future financial prospects. *Deal/Premoney valuation* is the ratio of the total amount of capital invested in the startup in the round to the startup's pre-money valuation in the same round.[23] A higher ratio signifies investors' commitment to supporting the startup for market expansion (Pham et al., 2023). *ln(Post valuation)* is the natural logarithm of the post valuation of the startup, which is the nominal value of the startup immediately after the VC financing round. A higher post valuation often suggests that investors see a significant market opportunity for the startup (Burt et al., 2023). *Generating revenue* is a dummy variable equal to one if the startup's business status is classified as "Generating Revenue" or "Profitable" by PitchBook, and zero otherwise. This variable captures the startup's success in translating its investments in climate technology into a revenue stream. For incumbents, the presence of a revenue generating startup serves as an indicator that real market demand exists for climate solutions.

In columns (1) to (3) of Table 4, we include triple interaction terms for *Deal/Premoney valuation*, *ln(Post valuation)* and *Generating revenue* interacted with *Top Similar* × *Post* in Equation (3), respectively. For all three columns, the coefficient on the triple interaction term is positive and statistically significant, implying that similar incumbent firms show a more pronounced increase in their *CS measure* in response to rounds with larger valuation and financial prospects.

4.3.2. Investor characteristics

We anticipate a stronger VC signal if the startup is funded by VCs that have a track record of superior investment performance and a commitment to investing in the climate-tech sector. VC performance is known to be persistent, with better performance in past investments predicting higher financial returns in subsequent investments (Nanda, Samila, & Sorenson, 2020). Therefore, incumbent firms may perceive the commitment of high-performing VCs in the climate-tech sector as a signal that active investment in climate-tech startups leads to higher financial returns (Burt et al., 2023).

To assess VCs' investment performance, we calculate the cash-on-cash (CoC) multiple of exited investments, the most common metric used by VCs to evaluate their performance (Gompers et al., 2020). The CoC multiple is the value of the startup at the time of exit divided by the total amount invested, representing the ratio of returned over invested capital

[23] The pre-money valuation of a startup is its valuation excluding the capital received in the latest financing round.

(Gaddy et al., 2017; van den Heuvel & Popp, 2023). For startups that are acquired, we use the reported deal acquisition value as the exit value. For an IPO exit, we use the pre-money IPO value as the exit value. Lastly, for startups exited through liquidation, bankruptcy, or going out of business, their exit value is set to zero. We assume a VC has a commitment to investments in the climate-tech sector if it has investment funds where "Climate Tech" or "CleanTech" is a preferred investment vertical according to PitchBook.

We define *High CoC VC* as a dummy variable equal to one if the average CoC multiple for all exited investments in the four years before the deal date across all VC investors in the syndicate is above the median and at least one VC investor in the syndicate is committed to climate-tech investments, and zero otherwise. In column (4) of Table 4, we introduce an interaction term between *High CoC VC* and *Top Similar* × *Post*. The positive and significant coefficient on this triple interaction term suggests that similar incumbent firms demonstrate a stronger positive response when startups are funded by high-performance VCs with a commitment to climate-tech investments, indicating a stronger signal for commercial potential.

We also test if incumbents exhibit a weaker response when impact investors participate in the funding round of a startup. Unlike traditional VCs, impact investors pursue non-financial objectives. Barber, Morse, and Yasuda (2021) show that impact investors are willing to sacrifice financial returns because they derive nonpecuniary utility from investments that yield beneficial environmental impact. Similar incumbent firms may interpret the presence of impact investors as an indication that the startup aligns more with broader environmental goals than with the commercial potential of climate solutions.

We account for the involvement of general impact investors using the variable *Impact investor*, which is a dummy variable equal to one if at least one of the investors participating in the VC financing round is classified as an impact investor by PitchBook, and zero otherwise. In column (5) of Table 4, the coefficient on the triple interaction term is negative and statistically significant, implying that the response of similar incumbent firms weakens when impact investors are part of the syndicate.[24]

[24] In Internet Appendix Table IA.1, we also consider two other types of strategic investors: Breakthrough Energy and corporate venture capital (CVC). While Breakthrough Energy operates as an impact investor with a broader focus than purely financial considerations, an investment from Breakthrough Energy likely signals both technical feasibility and the market scalability of the climate technology. CVCs may exhibit different risk appetites compared to traditional VCs, as they are more inclined to tolerate higher risks associated with innovative climate solutions (Chemmanur, Loutskina, & Tian, 2014; Ma, 2020). We find that the response of similar incumbent firms to startup investment by Breakthrough Energy or a CVC is not statistically different from that of a traditional VC.

4.3.3. Signaling visibility

Startup deals with higher visibility can indicate broader market validation for the commercial potential of climate technology offered by these startups. This enhanced visibility may boost the perceived market opportunities for the startup, thereby incentivizing incumbent firms to increase their focus on climate solutions.

We measure the visibility of startup deals in two ways. *New investors* is the number of new investors that participated in the VC financing round for the startup as a proportion of the total number of investors in the round. A new investor is someone who invests in a startup for the first time and has not participated in any prior round of financing for the same startup. We exclude seed rounds when calculating *New investors*.[25] An increase in the proportion of new investors in a given VC round signifies a broadening of the investor base, positively contributing to the visibility of the deal. *Media* is the natural logarithm of one plus the total number of news articles featuring the startup from four years prior to the event date to 30 days before the event date.[26] Research demonstrates that increased media coverage is associated with improved exit outcomes for startups (Baik & Shin, 2023). In columns (6) and (7) of Table 4, our findings indicate that VC rounds with higher visibility, as proxied by *New investors* and *Media*, correspond to a greater increase in similar incumbents' focus on climate solutions.

4.4. Endogeneity

Although the combination of the absence of pretrends and extensive cross-sectional analyses helps alleviate concerns about endogenous responses by similar incumbents to the VC signal, endogeneity issues may still remain. One concern is reverse causality—namely, that changes in similar incumbent firms' focus on climate solutions could influence the likelihood that climate-tech startups receive VC financing. Another concern is the presence of confounding shocks. For example, a technological breakthrough (e.g., improved materials for solar efficiency) or a policy intervention (e.g., new subsidies for solar manufacturing) could simultaneously drive both VC investment in startups and incumbents' increased focus on climate solutions. In such cases, the observed positive association between VC financing and incumbent response

[25] Seed rounds are excluded because in these rounds, all investors are typically regarded as new investors by default.

[26] PitchBook provides data on news articles related to a particular startup. We stop counting news articles 30 days before the deal date because, in the month leading up to the deal, there is often extensive news coverage of large deals. Thus, including these 30 days could potentially confound our results by capturing the large deal size effect documented earlier. Nonetheless, our findings remain qualitatively similar even if we include articles up until the deal date.

may reflect a correlated omitted variable rather than an exogenous response to the VC signal. In this section, we implement a series of tests to address these potential sources of endogeneity.

4.4.1. Instrumental variables using changes in state-level capital gains taxes

To explicitly address potential endogeneity concerns, we conduct a 2SLS regression using variation in state-level capital gains taxes faced by VC firms as an instrument for the probability that startups receive VC financing. This approach creates exogenous variation in the availability of the VC signal. In the U.S., VC firms are structured as "pass-through entities," meaning the firms themselves do not pay taxes; instead, profits are distributed to general partners (GPs), who pay taxes on these gains through their individual tax returns.[27] Since VC firms typically hold investments for the medium to long term, GPs are subject to capital gains taxes when they exit an investment by selling a capital asset at a profit.

Prior research shows that capital gains tax rates affect the investment behavior of GPs. A higher capital gains tax rate reduces GPs' after-tax returns and thus weakens their incentive to invest (Lerner & Nanda, 2020), prompting VC firms to reduce their supply of capital to startups (Keuschnigg & Nielsen, 2003, 2004). Consistent with this mechanism, Dimitrova and Eswar (2023) document that reductions in state-level capital gains taxes are associated with increased VC investment in startups. Accordingly, capital gains taxes are likely negatively correlated with the probability of startups receiving VC financing, satisfying the relevance condition for a valid instrument. At the same time, these taxes are unlikely to directly influence incumbent firms' focus on climate solutions, thereby satisfying the exclusion restriction.

Following Dimitrova and Eswar (2023), we obtain data on long-term capital gains tax rates from the NBER TAXSIM database and use the maximum state-level tax rate on long-term capital gains (*VC state tax*) as our instrumental variable.[28] Capital gains taxes are determined by the state of residence of the GPs, not the state of incorporation of the VC firm. Following prior work, we use the state in which the VC firm is headquartered to assign the applicable tax rate (Heider & Ljungqvist, 2015), assuming that GPs reside in the same state as the firm's headquarters (Lerner, 1995).[29]

To validate that changes in state-level capital gains taxes influence VC firms' investment behavior, we conduct panel regressions at the VC investor-year level, examining the impact of

[27] In the U.S., limited partners of VC firms are typically tax-exempt because they predominantly consist of pension funds and foundations (Lerner & Nanda, 2020).

[28] Since the actual tax rate for an individual is endogenous, the maximum state tax rate serves as a preferable instrument because it is exogenous to the labor supply and investment decisions of individuals.

[29] In cases where multiple VC investors participate in the syndicate financing a startup, we assign the tax rate based on the lead VC, as the lead VC typically plays the most active role in the deal.

VC state tax on VC investments in startups. In columns (1) and (2) of Panel A in Table 5, the dependent variable is a dummy equal to one if the VC investor finances a climate-tech startup in a given year, and zero otherwise (*VC investment*). We standardize *VC state tax* so that a one-unit change corresponds to a one standard deviation change, facilitating economic interpretation. The logit model in column (2) shows that a one standard deviation decrease in *VC state tax* increases the likelihood of VC investment in a climate-tech startup by 8.8 percentage points—an economically significant effect, representing roughly 20% of the standard deviation of *VC investment*. Columns (3) and (4) further show that lower *VC state tax* is associated with an increase in both the number and total size of climate-tech startup deals financed by the VC investor in a given year. Together, these results provide strong evidence of a negative correlation between capital gains taxes and VC firms' capital allocation to climate-tech startups.

To estimate the 2SLS regression, we begin by expressing our treatment variable as

$$\mathit{Top\ Similar}_i = \mathit{VC\ financed}_j \times 1(\mathit{Similarity\ score}_{j,i} \in \text{Top 1 percentile}) \tag{5}$$

where i denotes an incumbent firm, j denotes a startup, and t denotes year. *VC financed* is a dummy variable equal to one if the startup receives VC financing in a given year, and zero otherwise; $1(\cdot)$ is the indicator function; and *Similarity score* is defined in Equation (2). We use *VC state tax* as an instrument for *VC financed*, leading to the following first-stage regression:

$$\mathit{VC\ financed}_{j,t} = \beta_0 + \beta_1 \mathit{VC\ state\ tax}_{j,t} + \varepsilon_{j,t}. \tag{6}$$

We then take the predicted values of *VC financed*, denoted $\widehat{\mathit{VC\ financed}}$, to construct the predicted values of *Top Similar* as follows:

$$\widehat{\mathit{Top\ Similar}}_i = \widehat{\mathit{VC\ financed}}_j \times 1(\mathit{Similarity\ score}_{j,i} \in \text{Top 1 percentile}). \tag{7}$$

Finally, we use $\widehat{\mathit{Top\ Similar}}$ in the second-stage regression:

$$\mathit{CS\ measure}_{i,t,c} = \beta \widehat{\mathit{Top\ Similar}}_{i,c} \times \mathit{Post}_{t,c} + \gamma X_{i,t-1} + \tau_{i,c} + \rho_{t,c} + \varepsilon_{i,t,c}. \tag{8}$$

In Panel B of Table 5, we report the results from the 2SLS regressions. As expected, the first-stage results in column (1) show that a decrease in VC firms' state-level capital gains taxes significantly increases the probability that a startup receives VC financing.[30] The second-stage results, presented in columns (2) and (3), show that the coefficients on $\widehat{Top\ Similar} \times Post$ remain positive and statistically significant, with even larger economic magnitudes compared to the baseline estimates in Table 3. These findings help address endogeneity concerns by showing that the increase in similar incumbent firms' *CS measure* persists when the availability of VC investments in climate-tech startups is instrumented using exogenous variation from tax changes.

4.4.2. Confounding shocks

To assess whether our results are influenced by potential confounding shocks, we construct robust confidence intervals for the average treatment effect, accounting for possible violations of the parallel trends assumption, following Rambachan and Roth (2023). The intuition is that confounding shocks affecting similar incumbent firms relative to non-similar incumbents should create pre-treatment and post-treatment differences in trends in their *CS measure* that are unrelated to VC financing but instead driven by the confounding shock. The robust inference framework proposed by Rambachan and Roth (2023) enables us to bound the relative magnitude of post-treatment violations of parallel trends compared to the maximum pre-treatment violations of parallel trends. This approach allows us to evaluate whether the estimated average treatment effect remains statistically significant, even when deviations from the parallel trends assumption are permitted.

Figure 5 presents the robust 95% confidence intervals for the average treatment effect under increasingly larger deviations from the parallel trends assumption. The blue bar represents the original confidence interval for the average treatment effect, as reported in column (4) of Table 3, which assumes that the parallel trends assumption holds exactly. The figure shows that as deviations from parallel trends increase up to $\overline{M} = 1$—where post-treatment violations of parallel trends are restricted to be no larger than the maximal pre-treatment violation of parallel trends—the confidence intervals widen but remain above zero, allowing us to reject the null hypothesis of no effect. The average treatment effect becomes statistically insignificant only when $\overline{M} = 1.25$, indicating post-treatment violations exceed the maximum pre-treatment violations. Values of $\overline{M}$ exceeding 1 represent substantial departures from parallel trends, as Rambachan and Roth (2023) suggest $\overline{M} = 1$ as a reasonable benchmark when "the researcher suspects that possible violations of parallel trends are driven by confounding economic shocks

[30] The p-value for the Cragg and Donald (1993) instrument relevance test is less than 0.001, rejecting the null hypothesis that the instrument is weak.

that are of a similar magnitude to confounding economics shocks in the pre-period." Given that confounding factors, such as technological or policy developments, are unlikely to differ significantly between the pre- and post-treatment periods,[31] these results indicate that our baseline average treatment effect remains robust even under deviations from the parallel trends assumption.

To further rule out potential confounding shocks, we examine incumbents' responses to VC financing rounds, conditional on various climate and energy market variables. One concern is that shocks increasing attention to climate change could simultaneously influence both VC investment in climate-tech startups and similar incumbents' focus on climate solutions. To test for correlated timing effects, we use the *UMC* measure from Ardia et al. (2023), which captures unexpected changes in climate change concerns. In the first two columns of Table 6, we split the baseline analysis into periods of low and high *UMC* based on the median. If correlated timing were driving both VC financing and incumbents' responses, we would expect stronger effects in periods of high *UMC*, when general climate concerns and attention are more pronounced. However, the coefficient on *Top Similar* × *Post* is positive and significant in both subsamples, and the difference between subsamples is statistically insignificant, suggesting that the effects of VC financing are not disproportionately stronger during periods of heightened climate attention.

Next, we examine variables such as oil, natural gas, and solar prices to address potential confounding economic conditions. For example, periods of high oil or natural gas prices make recycled materials, heat pumps, or electric vehicles more economically attractive, potentially driving simultaneous increases in VC investments in climate-tech startups and similar incumbents' focus on climate solutions. Similarly, a technological shock that lowers solar prices might also affect incumbent and VC investment activity. In columns (3) to (8) of Table 6, we perform subsample analyses by splitting the data into periods of low and high oil, natural gas, and solar prices based on their respective medians. The coefficient on *Top Similar* × *Post* remains positive and significant across all subsamples, with no statistically significant differences between coefficients. These findings indicate that the documented association is unlikely to be driven by confounding shocks related to energy market conditions.

[31] For instance, widespread adoption of renewable energy technologies or phased implementation of subsidies for clean energy manufacturing often unfolds over multiple years, rather than abruptly, ensuring similar exposure across time. The four-year pre- and post-treatment windows represent a relatively short period during which these overarching technological or policy trends are unlikely to change significantly, reducing the likelihood of substantial divergence in confounding factors between the pre- and post-treatment periods.

4.4.3. *Mergers and acquisitions*

One alternative explanation for our findings is that the observed increase in *CS measure* could be due to incumbent firms acquiring climate-tech companies through M&A. In this scenario, a related concern is reverse causality, wherein M&A activity by incumbents could prompt VC investments in these startups. Another concern is that these incumbent firms might engage in killer acquisitions, where they purchase innovative startups with the intention of halting innovation projects to preempt future competition (Cunningham et al., 2021). To investigate these possibilities, we analyze incumbent firms' M&A activities involving startups before and after the VC deal dates. PitchBook provides data on the acquirers of startups that exited through M&A. For each of the incumbent firms in our sample, we use name matching to identify their involvement in the M&A of startups.[32]

In Figure 6, we conduct an event study analysis using the same specification as in Figure 4, except we replace *CS measure* with M&A activities. In Panel A, the outcome variable is a dummy variable that equals one if the incumbent firm acquires any startup (either climate-tech or non-climate-tech) in a given year, and zero otherwise. In Panel B, the outcome variable is a dummy variable that equals one if the incumbent firm acquires a climate-tech startup in a given year, and zero otherwise. In both panels, the coefficients are not statistically significantly different from zero in the years before or after VC financing rounds.

These findings alleviate concerns regarding reverse causality, as we do not detect significant differences in incumbents' acquisition of startups preceding VC financing rounds. Additionally, we do not see any significant increase in climate-tech acquisitions for similar incumbents in Panel B, which provides evidence against the killer acquisitions hypothesis and suggests that the increase in the *CS measure* is likely due to incumbent firms' internal development of climate solutions, rather than acquisitions of climate-tech startups.[33]

5. Additional analyses

5.1. *Which incumbent firms respond?*

The results thus far indicate that VC financing rounds for climate-tech startups prompt similar incumbent firms to intensify their focus on climate solutions. The cross-sectional tests by

[32]Specifically, we match each incumbent firm to the closest acquirer name through fuzzy matching in Stata. To ensure accuracy, we manually verify any matches with a match score below 100, which represents a perfect match.

[33]As we will show in Section 5.2, similar incumbent firms exhibit increases in both investment and research and development spending in response to the VC signal, which supports the interpretation that the rise in the *CS measure* reflects internal development efforts.

signaling strength suggest that these findings are consistent with the view that the VC signal enhances the perceived commercial potential of startups' climate technology, which in turn, motivates similar incumbent firms to focus on climate solutions. In this section, we corroborate these results by exploring which incumbent firms exhibit a more pronounced response to these VC signals.

If VC investment signals the commercial potential of climate solutions, then we expect two types of incumbent firms to be more likely to respond. First, we consider incumbent firms in the same industry as the startup. Incumbents and startups operating in the same industry are likely to share overlapping customer bases, which means the signal for customers' market demand is likely more relevant in the same industry. However, we acknowledge that incumbents in the same industry are also more likely to perceive the startup as a competitor in the same product space. As such, as a second prediction of the signaling effect, we expect incumbent firms with existing climate solutions to be more likely to respond to this signal because these firms possess complementary assets that facilitate the commercialization of climate solutions. In contrast, under the competitive threat channel, firms without existing climate solutions may be more likely to respond in order to prepare for future market threats.

We follow the same approach as in Section 4.3 by including triple interaction terms with these two types of incumbent firms. We manually match each startup's PitchBook industry code to at least one 4-digit GICS industry group based on the industry descriptions in PitchBook (following the procedure detailed in Section 2.2.1), and create an indicator, *Same industry*, that equals to one if the incumbent firm and the startup share the same 4-digit GICS industry group code, and zero otherwise. *Existing CS measure* is a dummy variable equal to one if the incumbent firm has at least one non-zero value of the *CS measure* in the four years leading up to the event date, and zero otherwise.

Consistent with our expectations, column (1) of Panel A in Table 7 shows that in VC rounds where similar incumbent firms operate in the same industry as the startup, there is a more pronounced increase in their focus on climate solutions. To ensure this effect is not solely driven by competition, column (2) demonstrates that similar incumbents with existing climate solutions exhibit a more substantial increase in their focus on climate solutions, indicating that the results are driven by those incumbents in a better position to capitalize on additional market demand.[34]

[34] To further distinguish between the VC signal channel and the competitive threat channel, we repeat our main analysis for the subset of startups in Seed and Series A rounds, as reported in Internet Appendix Table IA.2. The competitive threat from early-stage startups is likely weak, whereas VC investment in these early-stage startups remains a valuable signal for the commercial potential for climate solutions. We continue

5.1.1. *Stock price reaction*

As further corroborating evidence that the VC investments are events that provide information to the market that reflects the commercial potential for climate solutions, we examine the stock price reaction of firms that are more likely to benefit from this information. Given our findings that similar incumbents operating in the same industry as the startup and those with existing climate solutions exhibit stronger responses to VC investments, we hypothesize that these firms are more likely to benefit from the higher commercial potential for climate solutions. We assess how the shareholders of similar incumbent firms react to startups' VC financing rounds by analyzing the stock price changes around the deal date using a short-run event study methodology (MacKinlay, 1997). Studying equity market reactions allows us to infer shareholder expectations about the future benefits and costs associated with the responses of similar incumbent firms to these financing rounds.

We categorize the sample of similar incumbent firms across all VC financing rounds into four groups based on two criteria: whether the incumbent firm operates in the same industry as the startup and whether the incumbent firm has at least one non-zero value of the *CS measure* in the four years leading up to the event date. We calculate the 5-day $(-2, +2)$ and 11-day $(-5, +5)$ CARs around each deal date.[35] For benchmark returns, we estimate them using either the market model based on the CRSP value-weighted index, the four-factor Carhart (1997) model, or the 48 value-weighted industry return from Fama and French (1997).[36] We drop all incumbent firm-event date observations if the firm is identified as a similar firm in another round in the 30 days preceding and following the event date. This restriction ensures that each event consists only of firms affected by that singular event, reducing the spillover effects of stock price reactions to other nearby events.

Panel B of Table 7 presents the average CARs for each of the four groups of similar incumbent firms. The t-statistics for the mean (reported in the parenthesis) are calculated according to Boehmer, Musumeci, and Poulsen (1991) and account for event-induced changes in volatility. The results in columns (1) to (3) indicate that for similar incumbent firms operating in the same industry as the startup, those with prior climate solutions experience significantly higher CARs relative to those without. The positive effect on shareholder wealth is

to find a positive and statistically significant coefficient on *Top Similar* $\times$ *Post*, which is more consistent with the VC signal channel.

[35] To mitigate the impact of outliers, we apply winsorization to all CARs at the 1st and 99th percentiles.

[36] To estimate the benchmark model parameters for each firm-event date pair, we use 250 trading days of return data, with the window ending 20 days before the event date. We require a minimum of 120 non-missing observations within the estimation window.

also economically meaningful. Given that the average market capitalization of the incumbents in the same industry as the startup is approximately \$7.16 billion, the average difference in the $(-5, +5)$ CAR of 1.016% in column (3) using the market model translates to an estimated gain of approximately \$73 million (1.016% $\times$ \$7.16 billion) over the 11-day window. These results suggest that shareholders perceive similar incumbent firms with existing climate solutions and operating in the same industry as startups receiving VC financing as more likely to benefit from a greater commercial potential for climate solutions.

The results in columns (4) to (6) of Panel B of Table 7 portray a markedly different outcome for similar incumbent firms operating in different industries. Specifically, column (6) indicates that there are no statistically significant differences in the CARs between similar incumbents with and without existing climate solutions. The economic magnitude of the difference is also considerably smaller; for instance, the average difference in the $(-5, +5)$ CAR of 0.371% using the market model translates to a gain of only \$23 million. These findings suggest that there are no differential effects on shareholder valuation for similar incumbents with and without existing climate solutions when they operate in different industries from startups receiving VC financing.

5.2. Investment implications of incumbent firms' response

Given that similar incumbent firms' *CS measure* increases in response to VC investment, we examine whether such an increase translates to higher firm investments. However, the relationship between incumbents' *CS measure* and investments may be affected by endogeneity bias due to unobservable omitted variables that affect both incumbents' focus on climate solutions and their investment decisions. For example, firms with a strategic focus on innovation could be the same firms that focus more on climate solutions and, at the same time, make more investments.

To explicitly address potential endogeneity issues, we conduct 2SLS regressions using VC investment in climate-tech startups as an instrument for incumbents' *CS measure*. In the first stage, we regress *CS measure* on an instrument created based on the VC investment shock. In the second stage, we regress proxies for firm investment on the predicted value of *CS measure*. This approach ensures that we capture changes in firm investment driven by exogenous variation in *CS measure* instrumented by VC investments.

We use a staggered two-way fixed effects DiD specification for the first stage. Unlike stacked DiD, where we have to restrict the sample to create cohort-specific clean datasets, we do not need to discard any observations in a staggered DiD design. This design allows different incumbent firms to be treated at various points in time, allowing us to use the full time-series variation in our panel dataset. We estimate the first stage using the following firm-year panel regression with two-way fixed effects based on firm and year:

$$CS\ measure_{i,t} = \alpha + \beta\, Top\ Similar_{i,t-4:t} + \gamma X_{i,t-1} + \tau_i + \rho_t + \varepsilon_{i,t}, \tag{9}$$

where i denotes firm and t denotes year. The instrumental variable, $Top\ Similar_{i,t-4:t}$, is a dummy variable equal to one if the firm is identified as a similar firm of at least one startup during VC financing rounds in the current year or in the previous four years, and zero otherwise. $X_{i,t-1}$ includes the same control variables as in Equation (3). τ_i and ρ_t denote firm and year fixed effects, respectively.

In the second stage, we consider three proxies for firm investment:[37] 1) $\Delta CAPEX/Sales_{i,t+1}$ is the change in capital expenditures scaled by sales in years $t+1$ to t; 2) $\Delta R\&D/Sales_{i,t+1}$ is the change in research and development (R&D) expenses scaled by sales in years $t+1$ to t; and 3) $\Delta Div.\ payout/Assets_{i,t+1}$ is the change in total dividend payout scaled by total assets in years $t+1$ to t. We regress these proxies on the predicted values of *CS measure* from the first stage ($\widehat{CS\ measure}$).

Column (1) of Table 8 present the results from estimating Equation (9). Consistent with the baseline results using stacked DiD, the coefficient estimate of $Top\ Similar_{i,t-4:t}$ is positive and statistically significant, confirming the relevance condition of the instrumental variable. The p-value of the Cragg and Donald (1993) instrument relevance test is less than 0.001, rejecting the null hypothesis that the instrument is weak. Columns (2) to (4) present the results of the second stage. We find that an increase in instrumented *CS measure* leads to an increase in incumbent firms' change in capital expenditures and R&D expenses, both scaled by sales. The estimated coefficients are also economically significant. For example, a one-standard deviation increase in instrumented *CS measure* leads to an increase of 0.40 percentage points in the change in R&D expenses scaled by sales, corresponding to roughly 7% of its sample standard deviation. Column (4) shows that instrumented *CS measure* is negatively associated with changes in dividend payout scaled by total assets. This result is consistent with the view that incumbent firms with a higher *CS measure* reinvest funds internally towards capital expenditures and/or R&D to fund additional investments rather than distributing them to shareholders.

[37] We winsorize all three variables at the 1st and 99th percentiles to minimize the effect of outliers.

6. Robustness tests

6.1. *Alternative dependent variable for climate solutions*

We repeat our main analysis using an alternative measure for a firm's focus on climate solutions using the firm-level climate change exposure measures developed by Sautner et al. (2023). Their measure gauges the relative frequency with which bigrams related to climate change occur in the transcripts of earnings conference calls (*CCExposure*). Similar exposure variables are created to capture opportunities ($CCExposure^{Opp}$), regulatory shocks ($CCExposure^{Reg}$), and physical shocks ($CCExposure^{Phy}$) related to climate change.

In Panel A of Table 9, we use these firm-level exposure measures as the dependent variable in Equation (3). Column (1) shows that the overall exposure of similar incumbent firms to climate change increases following startups' VC financing rounds. A breakdown of the exposure into its three components reveals that this increased exposure is entirely due to additional climate-related opportunities. Specifically, the positive and significant coefficient on *Top Similar* $\times$ *Post* in column (2) implies that similar incumbent firms experience an increase of 0.008 percentage points in climate-related opportunities in response to startups' VC financing rounds, corresponding to 8% of the standard deviation of $CCExposure^{Opp}$. Columns (3) and (4) indicate that there are no significant effects on similar incumbent firms' exposure to regulatory and physical shocks related to climate change, respectively.

In Panel B of Table 9, we focus on opportunities related to climate change and examine other metrics besides exposure. Column (1) shows that VC financing rounds do not affect the uncertainties in climate-related opportunities for similar incumbent firms ($CCRisk^{Opp}$). Rather, these financing rounds induce an overall increase in the sentiment towards climate-related opportunities ($CCSentiment^{Opp}$), as shown in column (2). Specifically, column (3) documents an increase in positive sentiment related to these opportunities ($CCSentiment^{Opp,\ Pos}$), while column (4) shows there are no significant changes in negative sentiment ($CCSentiment^{Opp,\ Neg}$). Overall, the results in this section, employing the climate-related opportunity measure from Sautner et al. (2023), corroborate our primary findings.

6.2. *Alternative stacked DiD specifications*

We implement several alternative specifications for our baseline DiD regression. First, since not all incumbent firms may be involved in climate solutions, we restrict the sample of treated and control incumbent firms by including only those that have at least one non-zero value of *CS measure* during the event window. Second, to ensure that our results are not contingent on a specific choice of the control sample, we adopt alternative definitions. In one approach, we conduct a random one-to-one matching, where each treated firm in a given VC financing round is randomly matched to a firm sharing the same 6-digit GICS industry code but outside the top one percentile of similarity scores to serve as a control. The matched control firm cannot be identified as a similar firm for any other rounds in the entire event window. In another approach, we use different percentile cutoffs to define the pool of potential controls, such as the bottom 10, 30, and 40 percentiles of similarity scores. Internet Appendix Table IA.3 shows the results of these alternative specifications. In all columns, the coefficients on *Top Similar* × *Post* remain positive and statistically significant, indicating that our main results are not driven by firms with no *CS measure* during the event window and are robust across different control sample definitions.

6.3. Propensity score matching

One possible concern is that treated and control observations may not be directly comparable because they differ on other key dimensions. We use propensity score matching (PSM) to account for systematic differences between treated and control observations. The propensity score, $\widehat{p}$, is generated by estimating a logistic regression model, where the dependent variable is a dummy variable equal to one if the firm-year observation belongs to the treated group, and zero otherwise. The independent variables include all variables specified in the baseline model described in Equation (3).

For each treated firm in a given VC financing round, we match it to a firm sharing the same 6-digit GICS industry code with the closest propensity score (without replacement) outside the top one percentile of similarity scores to serve as a control (Roberts & Whited, 2013). The matched control firm cannot be identified as a similar firm for any other rounds in the entire event window. This matching procedure ensures that treated and control observations have similar propensity scores, accounting for systematic differences between the two groups. To assess the effectiveness of the matching procedure, Internet Appendix Table IA.4 shows that there are no observable differences between treated and control observations after the matching.

Using the matched sample, we re-estimate Equation (3), and the results are reported in columns (1) and (2) of Internet Appendix Table IA.5. The PSM results confirm our core finding that VC financing rounds prompt similar incumbent firms to increase their focus on climate solutions, reducing concerns that systematic differences between the treated and control groups drive our results.

Instead of discarding non-matched observations, an alternative approach is to incorporate all observations using a weighted least squares procedure. This method assigns weights that are inversely proportional to the probability of an observation being a treated or control unit. Specifically, we follow the procedure in Caliendo and Kopeinig (2008), whereby firm-year observations in the treated group receive a weight of $1/\widehat{p}$, while those in the control group receive a weight of $1/(1-\widehat{p})$. Intuitively, propensity score weighting assigns a lower weight to treated observations, which are "very different" (in terms of firm characteristics) from control observations and similarly, gives a lower weight to control observations, which are "very different" from treated observations. The results are presented in columns (3) and (4) of Internet Appendix Table IA.5. As before, the analysis demonstrates that VC financing rounds have a positive effect on similar incumbent firms' *CS measure*. Overall, the results in this section suggest that the relationship between VC financing rounds and *CS measure* is unlikely to be driven by selection bias.

6.4. Heterogeneous treatment effects

There are also concerns that heterogeneous treatment effects could yield biased estimates in DiD designs.[38] To address treatment effect heterogeneity, we estimate Equation (9) using the DiD estimator developed by de Chaisemartin and D'Haultfoeuille (2020).[39] The results, presented in columns (1) and (2) of Internet Appendix Table IA.6, show that our inferences continue to hold after controlling for treatment effect heterogeneity. Furthermore, none of the individual pre-trend estimators enter with statistically significant coefficients, and we fail to reject the null hypothesis that all pre-trend estimators equal zero. These analyses do not detect pre-trends in the four years before VC financing rounds after accounting for treatment effect heterogeneity.

7. Conclusion

Our study shows that VC investments in climate-tech startups can serve as an informative signal to validate the commercial potential of climate solutions, leading incumbent firms to increase their focus on climate solutions. Using stacked DiD, we find that incumbent firms operating in similar product markets as climate-tech startups receiving VC financing

[38] Heterogeneous treatment effects may occur because different subgroups of similar incumbent firms may react differently to a given VC financing round (heterogeneous treatment effects across groups) or a similar incumbent firm's response to latter VC financing rounds may be influenced by its response to earlier rounds (heterogeneous treatment effects across time).

[39] The de Chaisemartin and D'Haultfoeuille (2020) estimator is applicable in staggered DiD designs, as opposed to stacked DiD designs.

significantly increase their focus on climate solution products and services. In cross-sectional analysis, we observe that the increase is more pronounced when the VC signal is stronger, such as in VC rounds with a larger deal size, higher startup valuation, revenue-making startups, higher visibility, and when VCs are high-performing and financially motivated.

In response to the VC signal, we find that incumbent firms more likely to benefit from the signal increase their focus on climate solutions. Specifically, the increase in climate solutions is more pronounced for incumbent firms with pre-existing climate solutions and those in the same industry as the startup. Consistent with the VC investment signaling larger commercial potential for climate solutions, the stock price reactions around VC financing dates of similar incumbent firms operating in the same industry as the startup exhibit significantly higher CARs if they have prior climate solutions. Overall, our findings indicate that VC financing of climate-tech startups signals business opportunities that drive incumbents' focus on climate solutions and raise the possibility that VC investments might have an impact that extends beyond the direct effect on portfolio companies.

References

Acs, Z. J., & Audretsch, D. B. (1987). Innovation, market structure, and firm size. *Review of Economics and Statistics*, *69*(4), 567–574.

Aldatmaz, S., & Brown, G. W. (2020). Private equity in the global economy: Evidence on industry spillovers. *Journal of Corporate Finance*, *60*.

Allee, K. D., Christensen, T. E., Graden, B. S., & Merkley, K. J. (2021). The genesis of voluntary disclosure: An analysis of firms' first earnings guidance. *Management Science*, *67*(3), 1914–1938.

Ardia, D., Bluteau, K., Boudt, K., & Inghelbrecht, K. (2023). Climate change concerns and the performance of green vs. brown stocks. *Management Science*, *69*(12), 7151–7882.

Armstrong, C., Davila, A., & Foster, G. (2006). Venture-backed Private Equity Valuation and Financial Statement Information. *Review of Accounting Studies*, *11*, 119–154.

Azar, J., Duro, M., Kadach, I., & Ormazabal, G. (2021). The Big Three and corporate carbon emissions around the world. *Journal of Financial Economics*, *142*(2), 674–696.

Badertscher, B., Shroff, N., & White, H. D. (2013). Externalities of public firm presence: Evidence from private firms' investment decisions. *Journal of Financial Economics*, *109*(3), 682–706.

Baik, B. K., Berfeld, N., & Verdi, R. S. (2025). Do Public Financial Statements Influence Private Equity and Venture Capital Financing? *The Accounting Review*, *100*(2), 1–20.

Baik, B. K., & Shin, A. (2023). Breaking news! Venture capital influences startup media coverage. Harvard Business School Working Paper.

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395.

Barber, B. M., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Financial Economics*, *139*(1), 162–185.

Barrios, J. M., Choi, J. H., Hochberg, Y. V., Kim, J., & Liu, M. (2023). Informing Entrepreneurs? Initial Public Offerings and New Business Formation. University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-92.

Bernard, D., Kaya, D., & Wertz, J. (2021). Entry and capital structure mimicking in concentrated markets: The role of incumbents' financial disclosures. *Journal of Accounting and Economics*, *71*(2–3).

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. *Journal of Finance*, *71*(4), 1591–1622.

Bernstein, S., Lerner, J., Sorensen, M., & Strömberg, P. (2017). Private equity and industry performance. *Management Science*, *63*(4), 1198–1213.

BloombergNEF. (2021). Venture capital, PE invest $53.7 billion in climate tech. `https://www.bloomberg.com/professional/blog/venture-capital-pe-invest-53-7-billion-in-climate-tech/`.

Blyth, W., Bradley, R., Bunn, D., Clarke, C., Wilson, T., & Yang, M. (2007). Investment risks under uncertain climate change policy. *Energy Policy*, *35*(11), 5766–5773.

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

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, *142*(2), 517–549.

Bond, P., Edmans, A., & Goldstein, I. (2012). The real effects of financial markets. *Annual Review of Financial Economics*, *4*, 339–360.

Bourveau, T., Chen, J. V., Elfers, F., & Pierk, J. (2023). Public peers, accounting comparability, and value relevance of private firms' financial reporting. *Review of Accounting Studies*, *28*, 2642–2676.

Burt, A., Harford, J., Stanfield, J. R., & Zein, J. (2023). Does a VC's commitment lead to improved investment outcomes? Evidence from climate startups. UNSW Business School Research Paper.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

Cardinal, L. B., & Opler, T. C. (1995). Corporate diversification and innovative efficiency an empirical study. *Journal of Accounting and Economics*, *19*(2–3), 365–381.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *Quarterly Journal of Economics*, *134*(3), 1405–1454.

Chemmanur, T. J., Loutskina, E., & Tian, X. (2014). Corporate venture capital, value creation, and innovation. *Review of Financial Studies*, *27*(8), 2434–2473.

Chen, Q., Goldstein, I., & Jiang, W. (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies*, *20*(3), 619–650.

Cohen, L., Gurun, U. G., & Nguyen, Q. H. (2020). The ESG-innovation disconnect: Evidence from green patenting. NBER Working Paper No. 27990.

Cornelli, G., Frost, J., Gambacorta, L., & Merrouche, O. (2023). Climate tech 2.0: Social efficiency versus private returns. BIS Working Papers No. 1072.

Cragg, J. G., & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, *9*(2), 222–240.

Cunningham, C., Ederer, F., & Ma, S. (2021). Killer acquisitions. *Journal of Political Economy*, *129*(3), 649–702.

de Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–2996.

Dimitrova, L., & Eswar, S. K. (2023). Capital gains tax, venture capital, and innovation in start-ups. *Review of Finance*, *27*(4), 1471–1519.

Downar, B., Ernstberger, J., Reichelstein, S., Schwenen, S., & Zaklan, A. (2021). The impact of carbon disclosure mandates on emissions and financial operating performance. *Review of Accounting Studies*, *26*(3), 1137–1175.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, *43*(2), 153–193.

Ferracuti, E., & Stubben, S. R. (2019). The role of financial reporting in resolving uncertainty about corporate investment opportunities. *Journal of Accounting and Economics*, *68*(2–3).

Gaddy, B. E., Sivaram, V., Jones, T. B., & Wayman, L. (2017). Venture capital and cleantech: The wrong model for energy innovation. *Energy Policy*, *102*, 385–395.

Gans, J. S., Hsu, D. H., & Stern, S. (2002). When does start-up innovation spur the gale of creative destruction? *RAND Journal of Economics*, *33*(4), 571–586.

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*, *78*(2–3).

Goldstein, I., & Yang, L. (2019). Good disclosure, bad disclosure. *Journal of Financial Economics*, *131*(1), 118–138.

Gompers, P. A., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions? *Journal of Financial Economics*, *135*(1), 169–190.

Gormley, T. A., & Matsa, D. A. (2011). Growing out of trouble? corporate responses to liability risk. *Review of Financial Studies*, *24*(8), 2781–2821.

Gormley, T. A., & Matsa, D. A. (2014). Common errors: How to (and not to) control for

unobserved heterogeneity. *Review of Financial Studies*, *27*(2), 617–661.
Granskog, A., Patel, M., Gupta, R., & Helmcke, S. (2025). How incumbents can succeed in climate-driven growth investments. McKinsey Sustainability Report.
Hand, J. R. M. (2005). The Value Relevance of Financial Statements in the Venture Capital Market. *The Accounting Review*, *80*(2), 613–648.
Heider, F., & Ljungqvist, A. (2015). As certain as debt and taxes: Estimating the tax sensitivity of leverage from state tax changes. *Journal of Financial Economics*, *118*(3), 684–712.
Hellmann, T., & Puri, M. (2000). The interaction between product market and financing strategy: The role of venture capital. *Review of Financial Studies*, *13*(4), 959–984.
Hellmann, T., & Puri, M. (2002). Venture capital and the professionalization of start-up firms: Empirical evidence. *Journal of Finance*, *57*(1), 169–197.
Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *RAND Journal of Economics*, *24*(2), 248–270.
Henderson, R., & Serafeim, G. (2020). Tackling climate change requires organizational purpose. *American Economic Review: AEA Papers and Proceedings*, *110*, 177–180.
Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.
Hoberg, G., Phillips, G., & Prabhala, N. (2014). Product market threats, payouts, and financial flexibility. *Journal of Finance*, *69*(1), 293–324.
Howell, S. T., Lerner, J., Nanda, R., & Townsend, R. R. (2020). Financial distancing: How venture capital follows the economy down and curtails innovation. NBER Working Paper No. 27150.
Hsu, D. H. (2004). What do entrepreneurs pay for venture capital affiliation? *Journal of Finance*, *59*(4), 1805–1844.
IEA. (2024). Clean Energy Innovation. Energy Technology Perspectives Special Report on Clean Energy Innovation.
IRENA. (2024). Renewable power generation costs in 2023. International Renewable Energy Agency.
Kaplan, S. N., & Strömberg, P. (2001). Venture capitalists as principals: Contracting, screening, and monitoring. *American Economic Review: AEA Papers and Proceedings*, *91*(2), 426–430.
Kaplan, S. N., & Strömberg, P. (2004). Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *Journal of Finance*, *59*(5), 2177–2210.
Katz, S. P. (2009). Earnings Quality and Ownership Structure: The Role of Private Equity Sponsors. *The Accounting Review*, *84*(3), 623–658.
Keuschnigg, C., & Nielsen, S. B. (2003). Tax policy, venture capital, and entrepreneurship. *Journal of Public Economics*, *87*(1), 175–203.
Keuschnigg, C., & Nielsen, S. B. (2004). Start-ups, venture capitalists, and the capital gains tax. *Journal of Public Economics*, *88*(5), 1011–1042.
Kim, J., & Olbert, M. (2022). How does private firm disclosure affect demand for public firm equity? evidence from the global equity market. *Journal of Accounting and Economics*, *74*(2–3).
Kim, J., & Valentine, K. (2023). Public firm disclosures and the market for innovation. *Journal of Accounting and Economics*, *76*(1).
Kim, J., Verdi, R. S., & Yost, B. P. (2020). Do Firms Strategically Internalize Disclosure Spillovers? Evidence from Cash-Financed M&As. *Journal of Accounting Research*, *58*(5), 1249–1297.

Kortum, S., & Lerner, J. (2000). Assessing the contribution of venture capital to innovation. *RAND Journal of Economics*, *31*(4), 674–692.

Lerner, J. (1995). Venture capitalists and the oversight of private firms. *Journal of Finance*, *50*(1), 301–318.

Lerner, J., & 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), 237–261.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*, *66*(1), 35–65.

Lu, S., Riedl, E., Serafeim, G., & Xu, S. (2024). Climate solutions, transition risk, and stock returns. Harvard Business School Working Paper.

Lu, S., Serafeim, G., Xu, S., & Awada, M. (2024). Tracking business opportunities for climate solutions using AI in regulated accounting reports. Harvard Business School Working Paper.

Ma, S. (2020). The life cycle of corporate venture capital. *Review of Financial Studies*, *33*(1), 358–394.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, *35*(1), 13–39.

Martin, R., Muûls, M., & Wagner, U. J. (2016). The impact of the european union emissions trading scheme on regulated firms: what is the evidence after ten years? *Review of environmental economics and policy*.

Morsfield, S. G., & Tan, C. E. L. (2006). Do Venture Capitalists Influence the Decision to Manage Earnings in Initial Public Offerings? *The Accounting Review*, *81*(5), 1119–1150.

Naaraayanan, S. L., Sachdeva, K., & Sharma, V. (2021). The real effects of environmental activist investing. European Corporate Governance Institute - Finance Working Paper No. 743/2021.

Nanda, R., Samila, S., & Sorenson, O. (2020). The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics*, *137*(1), 231–248.

Noailly, J., Nowzohour, L., & van den Heuvel, M. (2022). Does environmental policy uncertainty hinder investments towards a low-carbon economy? NBER Working Paper No. 30361.

Noailly, J., & Smeets, R. (2015). Directing technical change from fossil-fuel to renewable energy innovation: An application using firm-level patent data. *Journal of Environmental Economics and Management*, *72*, 15–37.

Pham, P. K., Rezaei, R., & Zein, J. (2023). Venture capitalists vs. deep-pocketed incumbents: Startup financing strategies in the presence of competitive threats. UNSW Business School Research Paper.

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, *90*(5), 2555–2591.

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance. *Handbook of the Economics of Finance*, *2*, 493–572.

Samila, S., & Sorenson, O. (2011). Venture capital, entrepreneurship, and economic growth. *Review of Economics and Statistics*, *93*(1), 338–349.

Sautner, Z., van Lent, L., Vilkov, G., & Zhang, R. (2023). Firm-level climate change exposure. *Journal of Finance*, *78*(3), 1449–1498.

Schoonhoven, C. B., Eisenhardt, K. M., & Lyman, K. (1990). Speeding products to market: Waiting time to first product introduction in new firms. *Administrative Science Quarterly*, *35*(1), 177–207.

Shroff, N., Verdi, R. S., & Yost, B. P. (2017). When does the peer information environment matter? *Journal of Accounting and Economics*, *64*(2–3), 183–214.

Stuart, T. E., Hoang, H., & Hybels, R. C. (1999). Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, *44*(2), 315–349.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, *15*(6), 285–305.

Teixidó, J., Verde, S. F., & Nicolli, F. (2019). The impact of the eu emissions trading system on low-carbon technological change: The empirical evidence. *Ecological Economics*, *164*, 106347.

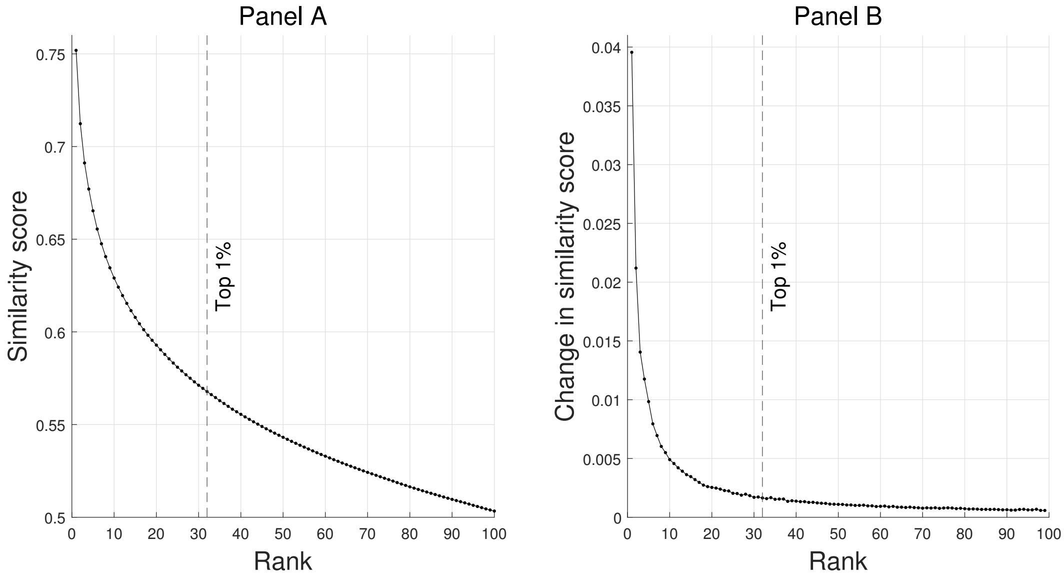
Tian, X., Udell, G. F., & Yu, X. (2016). Disciplining delegated monitors: When venture capitalists fail to prevent fraud by their IPO firms. *Journal of Accounting and Economics*, *61*(2–3), 526–544.

Tomar, S. (2023). Greenhouse gas disclosure and emissions benchmarking. *Journal of Accounting Research*, *61*(2), 451–492.

Tushman, M. L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, *31*(3), 439–465.

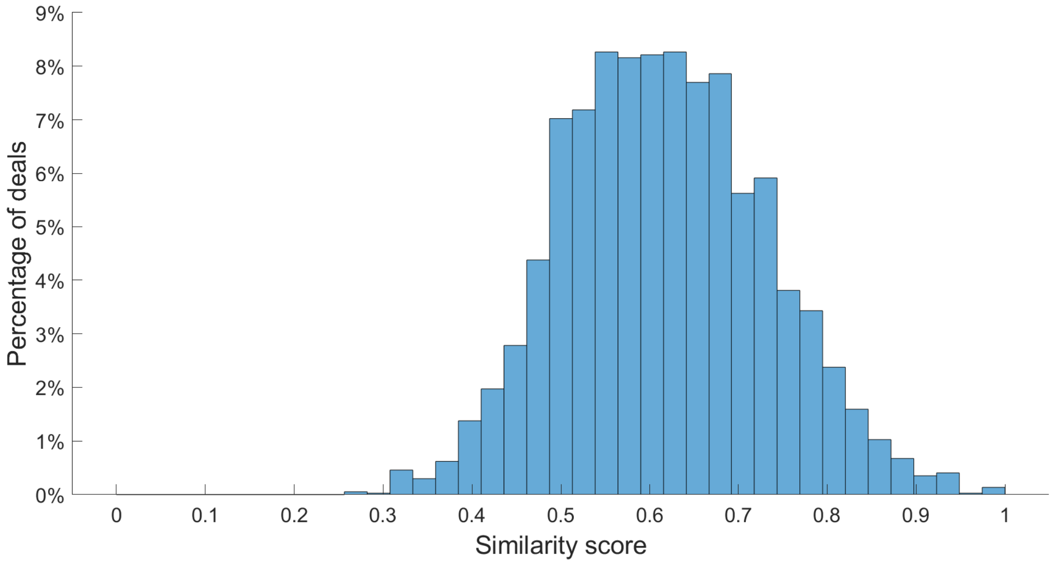
van den Heuvel, M., & Popp, D. (2023). The role of venture capital and governments in clean energy: Lessons from the first cleantech bubble. *Energy Economics*, *124*.

Figure 1
Ranking of incumbent firms based on similarity scores.



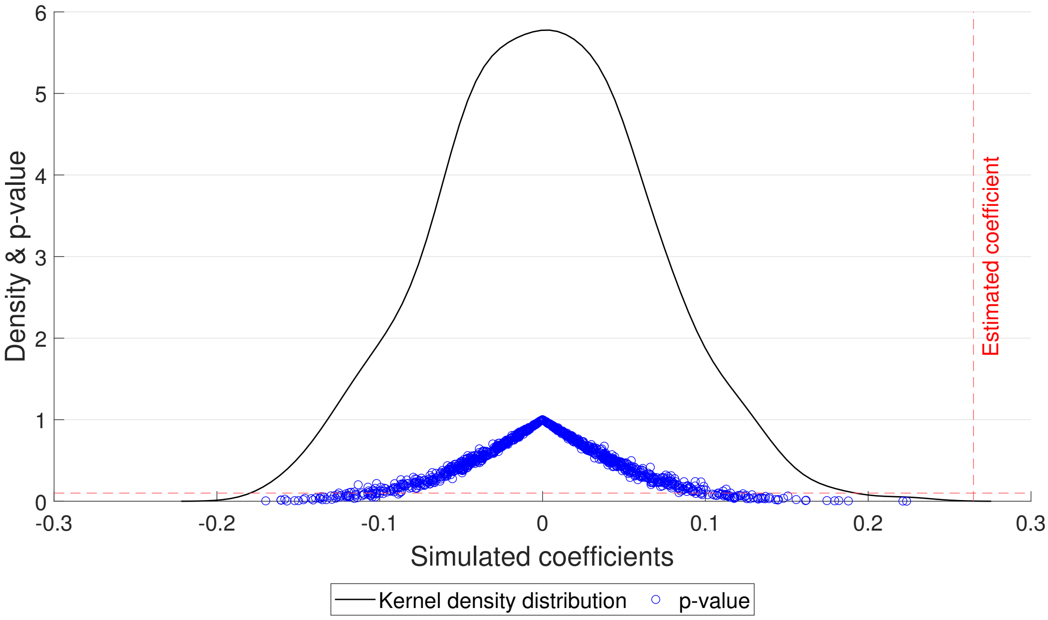
This figure shows how similarity scores vary across the 100 most similar incumbent firms to a given startup. For each VC financing round of a given startup, we compute a similarity score to capture the degree of overlap between the startup's and the incumbent firm's business descriptions. The horizontal axis is the rank of the incumbent, with the first rank indicating the incumbent firm with the highest similarity score to a given startup. In Panel A, the vertical axis is the average similarity score for a given rank across all VC financing rounds. In Panel B, the vertical axis is the change in the average similarity score between two consecutive ranks. The vertical dashed line represents the average cutoff rank for the top one percentile of similarity scores.

Figure 2
Distribution of average similarity scores among similar incumbent firms.



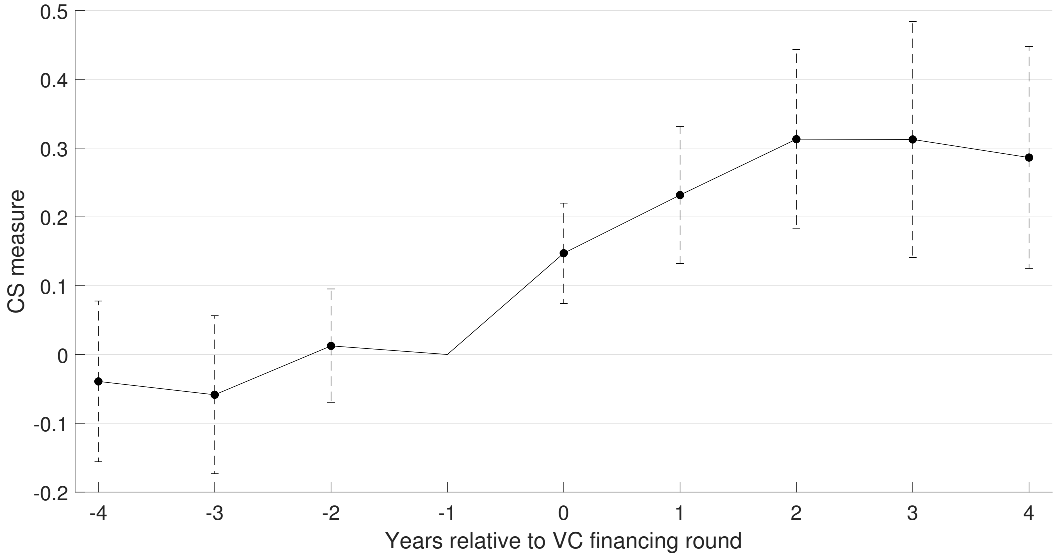
This figure shows the distribution of the average similarity scores between a startup and its top one percentile most similar incumbent firms across all VC financing rounds.

Figure 3
Placebo tests of stacked DiD estimates.



This figure presents placebo tests of the analysis in column (4) of Table 3. We estimate 1,000 simulations of the regression in column (4) of Table 3. In each simulation, we randomly assign *Top Similar* across incumbent firms rather than using the actual definition of *Top Similar*. We collect each simulation's estimated coefficient on *Top Similar* × *Post*. We then plot the kernel density distribution of these estimated coefficients and the corresponding p-values. The vertical dashed line is the coefficient on *Top Similar* × *Post* from column (4) of Table 3. The horizontal dashed line represents the 10% significance level.

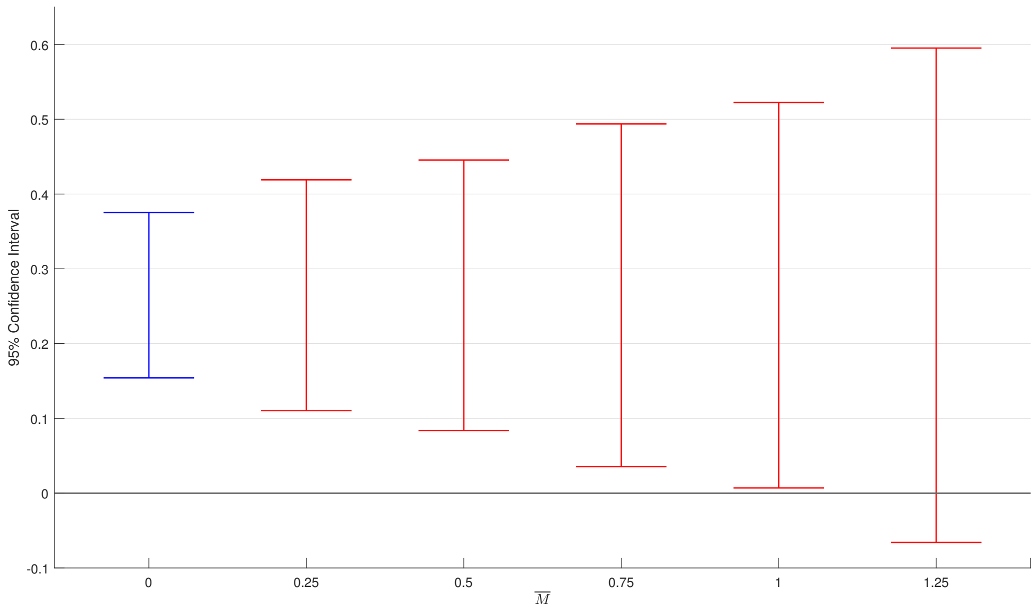
Figure 4
Dynamic stacked DiD estimates.



This figure plots the event study estimates and corresponding 95% confidence intervals according to the specification in Equation (4). We focus on an event window of four years before to four years after VC financing rounds. Event year $\ell = -1$ is the omitted category, implying that all coefficient estimates are relative to this year. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description.

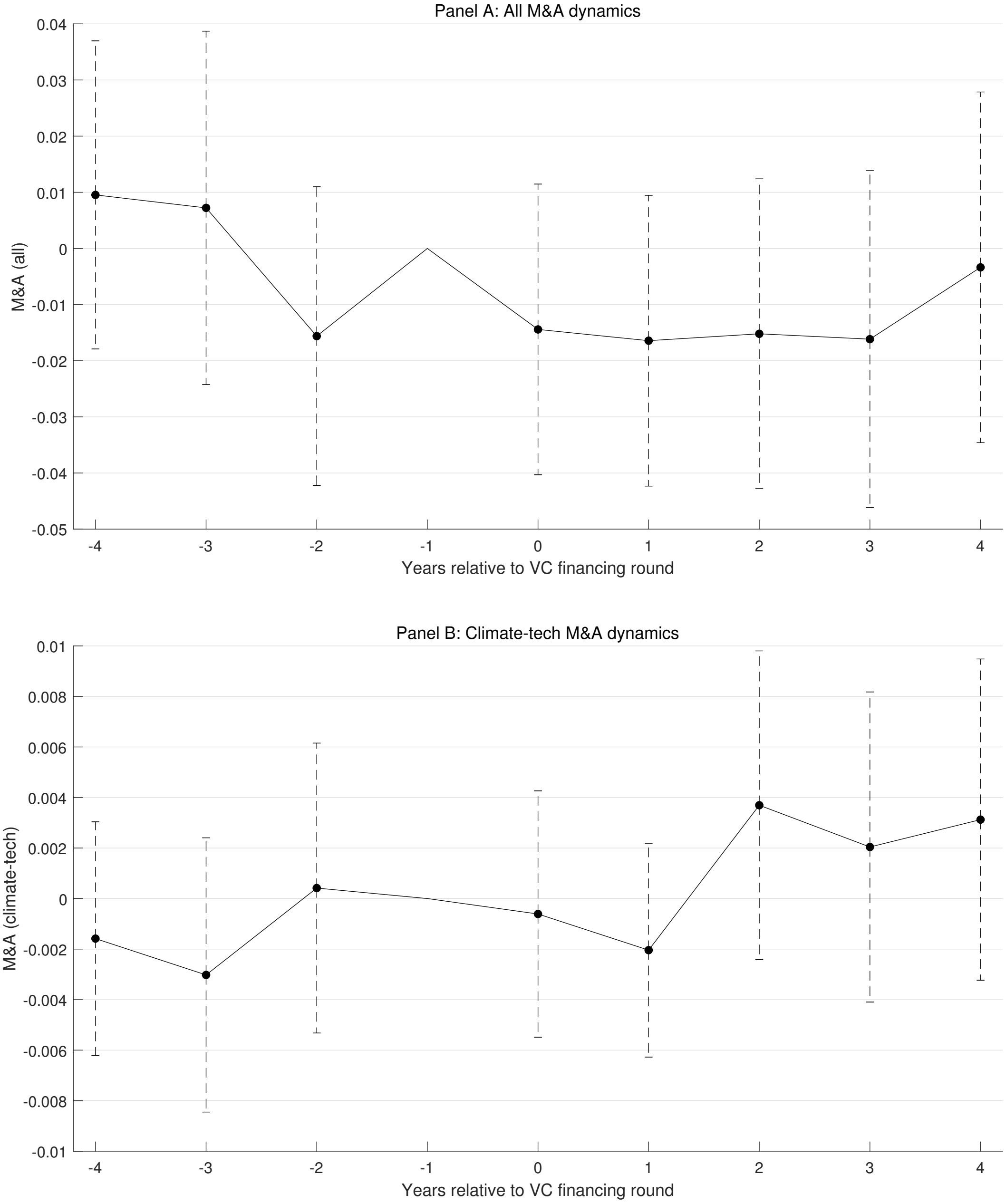
Figure 5

Robust confidence intervals of the average treatment effect accounting for potential violations of parallel trends.



This figure displays robust 95% confidence intervals for the average treatment effect across post-treatment periods, accounting for potential violations of the parallel trends assumption (Rambachan & Roth, 2023). The blue bar represents the original confidence interval for the average treatment effect, as reported in column (4) of Table 3. The red bars illustrate the robust confidence intervals for varying values of $\overline{M}$, which bound the relative magnitude of post-treatment violations of parallel trends compared to the maximum pre-treatment violations of parallel trends.

Figure 6
Dynamics of incumbent firms' M&A activity.



This figure plots the event study estimates and corresponding 95% confidence intervals for incumbents' M&A activity. The dependent variable in Panel A, M&A (all), is a dummy variable equal to one if the incumbent firm acquires a startup in a given year, and zero otherwise. The dependent variable in Panel B, M&A (climate-tech), is a dummy variable equal to one if the incumbent firm acquires a climate-tech startup in a given year, and zero otherwise. In both panels, we focus on an event window of four years before to four years after VC financing rounds. Event year $\ell = -1$ is the omitted category, implying that all coefficient estimates are relative to this year.

Table 1
Sample description.

Panel A: Climate-tech startup deals by year			
	Number of deals	Deal size ($M)	Number of investors per deal
Year	(1)	(2)	(3)
2005	54	7.770	3.3
2006	84	15.688	3.7
2007	166	15.248	3.3
2008	216	17.649	3.0
2009	163	16.076	3.0
2010	203	18.086	3.2
2011	229	15.593	3.2
2012	213	17.513	3.5
2013	215	9.427	2.9
2014	219	11.522	4.2
2015	187	14.114	4.0
2016	218	20.557	3.7
2017	233	14.189	3.9
2018	290	22.937	4.2
2019	304	15.984	4.8
2020	389	25.469	4.8
2021	578	40.740	5.8

Panel B: Startups' industry distribution		
PitchBook industry sector	Number of startups	Percent
(1) Business Products and Services	595	31.04
(2) Consumer Products and Services	224	11.68
(3) Energy	492	25.67
(4) Financial Services	19	0.99
(5) Healthcare	31	1.62
(6) Information Technology	318	16.59
(7) Materials and Resources	238	12.42
Total	1,917	100.00

Panel C: Similar incumbent firms' industry distribution		
GICS industry sector	Number of incumbents	Percent
(10) Energy	189	15.76
(15) Materials	136	11.34
(20) Industrials	318	26.52
(25) Consumer discretionary	135	11.26
(30) Consumer staples	110	9.17
(45) Information technology	266	22.19
(55) Utilities	45	3.75
Total	1,199	100.00

This table provides the description of our sample over the period 2005–2021. Panel A summarizes all climate-tech startup deals with VC financing by year. "Number of deals" is a count of all VC round-level investments made in climate-tech startups each year. "Deal size ($M)" is the average size of the investment in millions of dollars. "Number of investors per deal" is the average number of investors that take part in a round of financing. Panel B presents the industry distribution of climate-tech startups based on PitchBook's industry sector. Panel C presents the industry distribution of similar incumbent firms based on GICS industry sector. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup.

Table 2
Summary statistics of variables.

Variables	N	Mean	Median	P25	P75	Std. dev.
Dependent variables						
CS measure	59,898	0.724	0.000	0.000	0.299	2.494
Δ*CAPEX/Sales*	21,142	-0.018	0.000	-0.009	0.009	0.167
Δ*R&D/Sales*	21,142	-0.001	0.000	-0.000	0.001	0.062
Δ*Div. payout/Assets*	19,683	0.000	0.000	-0.004	0.005	0.073
VC investment	24,626	0.230	0.000	0.000	0.000	0.421
Number of deals	24,626	0.365	0.000	0.000	0.000	0.977
Deal size	24,626	14.720	0.000	0.000	0.000	130.013
CCExposure	40,071	0.105	0.049	0.019	0.113	0.192
$CCExposure^{Opp}$	40,071	0.029	0.000	0.000	0.022	0.095
$CCExposure^{Reg}$	40,071	0.005	0.000	0.000	0.000	0.018
$CCExposure^{Phy}$	40,071	0.002	0.000	0.000	0.000	0.012
$CCRisk^{Opp}$	40,071	0.001	0.000	0.000	0.000	0.008
$CCSentiment^{Opp}$	40,071	0.012	0.000	0.000	0.006	0.042
$CCSentiment^{Opp,\ Pos}$	40,071	-0.005	0.000	0.000	0.000	0.023
$CCSentiment^{Opp,\ Neg}$	40,071	0.007	0.000	0.000	0.000	0.034
Explanatory variables						
$Top\ Similar_{i,c}$	59,898	0.161	0.000	0.000	0.000	0.368
$Top\ Similar_{i,t-4:t}$	22,345	0.632	1.000	0.000	1.000	0.482
$CS\ \widehat{measure}$	22,345	2.184	2.199	1.981	2.397	0.310
Deal/Premoney valuation	30,260	0.548	0.417	0.224	0.684	0.538
ln(Post valuation)	30,282	3.216	3.102	2.303	3.993	1.378
Generating revenue	59,898	0.609	1.000	0.000	1.000	0.488
High CoC VC	32,039	0.109	0.000	0.000	0.000	0.311
Impact investor	59,898	0.308	0.000	0.000	1.000	0.462
New investors	28,663	0.737	0.800	0.500	1.000	0.270
Media	59,898	0.268	0.000	0.000	0.000	0.540
Same industry	59,898	0.358	0.000	0.000	1.000	0.479
Existing CS measure	59,898	0.263	0.000	0.000	1.000	0.440
VC state tax	24,626	1.681	1.504	1.080	2.902	1.000
Firm characteristics						
Firm age	59,898	3.092	3.258	2.944	3.401	0.483
Firm size	59,898	6.223	5.986	4.441	7.723	2.278
Book-to-market	59,898	0.541	0.538	0.408	0.662	0.195
ROA	59,898	0.086	0.109	0.055	0.167	0.192
Leverage	59,898	0.182	0.122	0.011	0.278	0.197
Sales growth	59,898	0.151	0.050	-0.057	0.178	2.661
Cash	59,898	0.138	0.084	0.027	0.192	0.157
Momentum	59,898	1.125	1.041	0.820	1.304	0.745
Stock return	59,898	0.131	0.057	-0.200	0.308	0.728
Stock volatility	59,898	0.123	0.103	0.073	0.149	0.079
Market share	59,898	0.019	0.002	0.000	0.009	0.058
Fluidity	59,898	4.224	3.459	2.288	5.412	2.731

This table reports summary statistics for the variables used in our analysis for the sample period from fiscal year 2005 to 2021. Std. dev. displays the standard deviation, P25 the first and P75 the third quartile of the respective variable. Variable definitions are presented in Table A.1 in Appendix A.

Table 3
Baseline stacked DiD estimates.

Dep. variable: *CS measure*	(1)	(2)	(3)	(4)
Top Similar × *Post*	0.205***	0.205***	0.210***	0.265***
	(4.06)	(4.10)	(4.03)	(4.70)
Firm age		0.090	0.224	0.363*
		(0.50)	(1.21)	(1.66)
Firm size		0.055	0.054	0.060
		(1.05)	(0.80)	(0.94)
Book-to-market		-0.030	0.027	0.187
		(-0.19)	(0.14)	(1.24)
ROA		-0.018	-0.080	-0.081
		(-1.11)	(-0.82)	(-0.90)
Leverage		-0.410**	-0.418**	-0.486**
		(-2.23)	(-2.10)	(-2.21)
Sales growth		0.001	0.000	-0.001
		(1.26)	(1.22)	(-0.25)
Cash		0.163	0.221	0.233
		(0.74)	(0.95)	(0.98)
Momentum			0.006	0.012
			(0.52)	(0.93)
Stock return			-0.042	-0.047
			(-1.19)	(-1.32)
Stock volatility			0.141	0.135
			(0.57)	(0.58)
Market share				-0.481
				(-0.43)
Fluidity				0.008
				(0.92)
Firm × Cohort F.E.	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes
Observations	73,019	68,041	63,367	59,898
Adj R^2	0.90	0.89	0.89	0.90

This table reports results from firm-level stacked DiD regressions examining the effect of climate-tech startup deals on incumbent firms' climate solutions measure. We focus on an event window of four years before to four years after VC financing rounds. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. An incumbent firm is treated in a given VC financing round if it is identified as a similar firm in that round and not in any other rounds in the preceding four years. If a treated firm is classified as a similar firm in multiple startup deals that occur in the same event year, we assign it to the deal where it has the highest similarity score. An incumbent firm is a control in a given VC financing round if it is in the bottom 20 percentile of similarity scores and is not identified as a similar firm for any other rounds in the entire event window. Control firms are matched to the same 6-digit GICS industry code as the treated firms. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 4
Heterogeneity across startup deals.

Dep. variable: *CS measure*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Top Similar × *Post*	0.173***	0.021	0.127***	0.178***	0.310***	-0.151	0.206***
	(3.28)	(0.22)	(2.82)	(3.16)	(7.73)	(-1.27)	(5.20)
Top Similar × *Post* × *Deal/Premoney valuation*	0.133***						
	(2.83)						
Top Similar × *Post* × *ln(Post valuation)*		0.072**					
		(2.45)					
Top Similar × *Post* × *Generating revenue*			0.237***				
			(4.05)				
Top Similar × *Post* × *High CoC VC*				0.310***			
				(2.56)			
Top Similar × *Post* × *Impact investor*					-0.129**		
					(-2.29)		
Top Similar × *Post* × *New investors*						0.499***	
						(3.26)	
Top Similar × *Post* × *Media*							0.235***
							(3.10)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,260	30,282	59,898	32,039	59,898	28,663	59,898
Adj R^2	0.92	0.92	0.91	0.92	0.90	0.92	0.91

This table reports results from firm-level stacked DiD regressions examining the cross-sectional heterogeneity across startup deals. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. *Deal/Premoney valuation* is the ratio of the total amount of capital invested in the startup in the round to the pre-money valuation of the startup. *ln(Post valuation)* is the natural logarithm of the post valuation of the startup. *Generating revenue* is a dummy variable equal to one if the startup's business status is classified as "Generating Revenue" or "Profitable", and zero otherwise. *High CoC VC* is a dummy variable equal to one if the average cash-on-cash multiple in the four years before the deal date across all VC investors in the syndicate is above the median and at least one VC investor in the syndicate has investment funds where "Climate Tech" or "CleanTech" is a preferred investment vertical according to Pitchbook, and zero otherwise. *Impact investor* is a dummy variable equal to one if at least one of the investors participating in the VC financing round is classified as an impact investor by PitchBook, and zero otherwise. *New investors* is the number of new investors that participated in the VC financing round for the startup as a proportion of the total number of investors in the round. A new investor is someone who invests in a startup for the first time and has not participated in any prior round of financing for the same startup. We exclude seed rounds when calculating *New investors*. *Media* is the natural logarithm of one plus the total number of news articles featuring the startup from four years prior to the event date to 30 days before the event date. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity.* For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 5
Instrumental variables using changes in state-level capital gains taxes.

Panel A: State-level capital gains taxes and VC investments

Dep. variable:	*VC investment*		$ln(1 + Number\ of\ deals)$	$ln(1 + Deal\ size)$
	LPM	Logit		
	(1)	(2)	(3)	(4)
VC state tax	-0.053**	-0.374**	-0.059**	-0.133**
	(-2.55)	(-2.55)	(-2.55)	(-2.10)
VC investor F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Observations	24,626	24,212	24,626	24,626
Adj R^2	0.14	0.02	0.25	0.21

Panel B: 2SLS estimation

	First stage	Second stage	
Dep. variable:	*VC financed*	*CS measure*	
	(1)	(2)	(3)
VC state tax	-0.445***		
	(-2.72)		
$\widehat{Top\ Similar} \times Post$		0.245***	0.328***
		(3.35)	(4.40)
Controls	No	No	Yes
Startup F.E.	Yes	No	No
Year F.E.	Yes	No	No
Firm × Cohort F.E.	No	Yes	Yes
Year × Cohort F.E.	No	Yes	Yes
Observations	6,137	69,001	56,462
Adj R^2	0.01	0.89	0.91

This table reports results using changes in state-level capital gains taxes as instrumental variables. Panel A reports regressions at the VC investor–year level. The independent variable, *VC state tax*, is the maximum state-level long-term capital gains tax rate in the headquarter state of the VC investor, normalized such that a unit change corresponds to one standard deviation for interpretability. The dependent variable in columns (1) and (2) is a dummy variable equal to one if the VC investor finances a climate-tech startup in a given year, and zero otherwise (*VC investment*). Column (1) uses a linear probability model and column (2) uses a logit model. The dependent variables in columns (3) and (4) are the natural logarithm of one plus the total number of climate-tech startups that the VC investor finances in a given year ($ln(1 + Number\ of\ deals)$) and the natural logarithm of one plus the total size of climate-tech startup deals that the VC investor finances in a given year ($ln(1 + Deal\ size)$), respectively. Panel B estimates a 2SLS model. The first stage is estimated at the startup–year level, where the dependent variable is an indicator equal to one if the startup receives VC financing in a given year , and zero otherwise (*VC financed*). In this stage, we use *VC state tax* as the instrument without normalization. In the second stage, we estimate firm-level stacked DiD regressions, using the instrumented treatment variable, $\widehat{Top\ Similar}$, as defined in Equation (7). The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity*. For all specifications, standard errors are robust to heteroskedasticity and are clustered at the state-level in Panel A, at the startup-level in column (1) of Panel B, and at the firm-level in columns (2) and (3) of Panel B; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 6
Incumbents' response split by climate and energy market variables.

	UMC		*Oil prices*		*NG prices*		*Solar prices*	
	Low	High	Low	High	Low	High	Low	High
Dep. variable: *CS measure*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Top Similar × *Post*	0.218***	0.269***	0.319***	0.215***	0.259***	0.272***	0.316***	0.225***
	(2.87)	(3.20)	(3.59)	(3.18)	(3.32)	(3.67)	(3.12)	(3.24)
Coefficient difference	0.052		-0.104		0.013		-0.091	
t-statistic	(0.49)		(-0.95)		(0.13)		(-0.71)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,944	23,509	27,236	32,662	30,958	28,940	25,646	34,252
Adj R^2	0.92	0.89	0.89	0.92	0.89	0.92	0.89	0.92

This table examines incumbents' response to VC financing rounds conditional on various climate and energy market variables. We focus on an event window of four years before to four years after VC financing rounds. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. An incumbent firm is treated in a given VC financing round if it is identified as a similar firm in that round and not in any other rounds in the preceding four years. If a treated firm is classified as a similar firm in multiple startup deals that occur in the same event year, we assign it to the deal where it has the highest similarity score. An incumbent firm is a control in a given VC financing round if it is in the bottom 20 percentile of similarity scores and is not identified as a similar firm for any other rounds in the entire event window. Control firms are matched to the same 6-digit GICS industry code as the treated firms. Columns (1) and (2) report the results for the subsample where *UMC* is below ("Low") or above ("High") the median, respectively. *UMC* is the unexpected climate change concerns measure of Ardia et al. (2023) in the month of the deal. Columns (3) and (4) report the results for the subsample where the annual crude oil prices are below ("Low") or above ("High") the median, respectively. Crude oil prices are measured in the year of the deal, with data obtained from the St. Louis Fed. Columns (5) and (6) report the results for the subsample where the annual natural gas prices are below ("Low") or above ("High") the median, respectively. Natural gas prices are measured in the year of the deal, with data obtained from the St. Louis Fed. Columns (7) and (8) report the results for the subsample where the solar photovoltaic module price is below ("Low") or above ("High") the median, respectively. Prices are measured in the year of the deal, with data obtained from IRENA (2024). Coefficient difference represents the difference in the coefficient estimates of *Top Similar* × *Post* between the two subsamples. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity*. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 7
Incumbents' response and stock price reaction split by industry and existing climate solutions.

Panel A: Incumbents' response		
Dep. variable: *CS measure*	(1)	(2)
Top Similar × *Post*	0.203***	0.144***
	(5.31)	(2.57)
Top Similar × *Post* × *Same industry*	0.143**	
	(2.20)	
Top Similar × *Post* × *Existing CS measure*		0.268***
		(2.67)
Controls	Yes	Yes
Firm × Cohort F.E.	Yes	Yes
Year × Cohort F.E.	Yes	Yes
Observations	59,898	59,898
Adj R^2	0.91	0.90

Panel B: Stock price reaction of similar incumbents						
	Same industry			Different industry		
	Existing CS measure ($N = 1,124$)	No *Existing CS measure* ($N = 1,378$)	Difference: (1) − (2)	*Existing CS measure* ($N = 2,822$)	No *Existing CS measure* ($N = 5,516$)	Difference: (4) − (5)
	(1)	(2)	(3)	(4)	(5)	(6)
Market model						
$(-2, +2)$	0.392**	-0.139	0.532**	0.174	0.163	0.011
	(2.11)	(-0.83)	(2.15)	(1.48)	(1.11)	(0.06)
$(-5, +5)$	0.507*	-0.509*	1.016**	0.423*	0.053	0.371
	(1.76)	(-1.81)	(2.54)	(1.92)	(0.27)	(1.26)
Four-factor model						
$(-2, +2)$	0.322*	-0.190	0.512**	0.131	0.137	-0.007
	(1.79)	(-1.12)	(2.10)	(1.12)	(0.94)	(-0.04)
$(-5, +5)$	0.606**	-0.467	1.073***	0.243	-0.080	0.322
	(2.11)	(-1.63)	(2.65)	(1.10)	(-0.42)	(1.10)
Industry model						
$(-2, +2)$	0.384**	-0.156	0.540**	0.180	0.188	-0.007
	(2.06)	(-0.93)	(2.17)	(1.54)	(1.27)	(-0.04)
$(-5, +5)$	0.555*	-0.507*	1.062***	0.406*	0.080	0.327
	(1.92)	(-1.79)	(2.65)	(1.85)	(0.41)	(1.11)

This table examines incumbents' response to VC financing rounds and their stock price reaction to these events conditional on whether the incumbent operates in the same industry as the startup and whether the incumbent has existing climate solutions. In Panel A, we estimate firm-level stacked DiD regressions by including additional triple interaction terms. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. *Same industry* is a dummy variable equal to one if the incumbent firm and the startup share the same 4-digit GICS industry group code, and zero otherwise. We manually match each startup's PitchBook industry code to at least one 4-digit GICS industry group based on the descriptions provided by the respective industry taxonomy. *Existing CS measure* is a dummy variable equal to one if the incumbent firm has at least one non-zero value of *CS measure* in the four years leading up to the event date, and zero otherwise. In Panel B, we report the mean CARs (in %) of similar incumbent firms around VC financing dates split based on *Same industry* and *Existing CS measure*. We consider event windows of 5 $(-2, +2)$ and 11 $(-5, +5)$ days. If an incumbent firm is classified as a similar firm in multiple startup deals that occur in the same event year, we assign it to the deal where it has the highest similarity score. We drop all incumbent firm–event date observations if the firm is identified as a similar firm in another round in the 30 days preceding and following the event date. CARs are risk adjusted using the market model, Carhart's (1997) four-factor model, and Fama and French's (1997) 48 value-weighted industry return. The t-statistics for the mean (reported in the parenthesis) account for event-induced changes in volatility and are calculated according to Boehmer et al. (1991). *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 8
Climate solutions and firm investments.

	First stage	Second stage		
Dep. variable:	$CS\ measure_{i,t}$	$\Delta CAPEX/Sales_{i,t+1}$	$\Delta R\&D/Sales_{i,t+1}$	$\Delta Div.\ payout/Assets_{i,t+1}$
	(1)	(2)	(3)	(4)
$Top\ Similar_{i,t-4:t}$	0.176***			
	(3.07)			
$\widehat{CS\ measure}_{i,t}$		0.038*	0.013**	-0.017*
		(1.87)	(1.97)	(-1.84)
Firm age	-0.285	0.064***	0.001	0.006
	(-1.11)	(4.18)	(0.11)	(1.04)
Firm size	0.145*	-0.032***	0.000	0.015***
	(1.84)	(-5.21)	(0.17)	(5.16)
Book-to-market	-0.032	-0.006	0.012	0.045***
	(-0.16)	(-0.26)	(1.45)	(4.99)
ROA	-0.015	0.088***	0.042**	0.023***
	(-0.83)	(3.40)	(2.53)	(9.44)
Leverage	-0.445**	-0.006	0.005	-0.025***
	(-2.20)	(-0.31)	(0.80)	(-2.63)
Sales growth	0.005	0.004**	0.003***	0.001
	(1.40)	(2.79)	(10.01)	(1.10)
Cash	-0.081	0.050*	0.010	0.076***
	(-0.29)	(2.12)	(0.76)	(5.77)
Momentum	0.013	-0.001	0.000	0.002
	(0.42)	(-0.28)	(0.00)	(1.61)
Stock return	-0.080**	0.010***	0.003*	0.001
	(-2.40)	(2.95)	(1.71)	(0.47)
Stock volatility	0.286	0.018	0.006	0.013
	(0.79)	(0.47)	(0.38)	(0.88)
Market share	0.062	0.133**	-0.008	-0.029
	(0.09)	(2.49)	(-0.61)	(-0.91)
Fluidity	-0.015	0.000	0.001	-0.001**
	(-1.20)	(0.19)	(1.50)	(-2.46)
Cragg-Donald test	(p-value < 0.001)			
Firm F.E.	Yes	Yes	Yes	Yes
Year F.E.	Yes	Yes	Yes	Yes
Observations	22,345	21,142	21,142	19,683
Adj R^2	0.93	0.14	0.07	0.10

This table reports results from 2SLS regressions examining the relationship between incumbent firms' *CS measure* and investments. The unit of observation is an incumbent firm–year. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. Column (1) estimates the first stage staggered DiD regression in Equation (9) using standard two-way fixed effects. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. $Top\ Similar_{i,t-4:t}$ is a dummy variable equal to one if the firm is identified as a similar firm of at least one startup during VC financing rounds in the current year or in the previous four years, and zero otherwise. Columns (2) to (4) regress proxies for firm investment on the predicted values of *CS measure* from the first stage ($\widehat{CS\ measure}$). $\Delta CAPEX/Sales_{i,t+1}$ is the change in capital expenditures scaled by sales in years $t+1$ to t. $\Delta R\&D/Sales_{i,t+1}$ is the change in research and development expenses scaled by sales in years $t+1$ to t. $\Delta Div.\ payout/Assets_{i,t+1}$ is the change in total dividend payout scaled by total assets in years $t+1$ to t. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table 9
Validation using Sautner et al.'s (2023) measure.

Panel A: Exposure measures				
Dep. variable:	*CCExposure*	$CCExposure^{Opp}$	$CCExposure^{Reg}$	$CCExposure^{Phy}$
	(1)	(2)	(3)	(4)
Top Similar × *Post*	0.010**	0.008***	-0.000	-0.000
	(2.36)	(3.07)	(-0.49)	(-0.47)
Controls	Yes	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes
Observations	39,148	39,148	39,148	39,148
Adj R^2	0.80	0.78	0.44	0.51
Panel B: Opportunity measures				
Dep. variable:	$CCRisk^{Opp}$	$CCSentiment^{Opp}$	$CCSentiment^{Opp,\ Pos}$	$CCSentiment^{Opp,\ Neg}$
	(1)	(2)	(3)	(4)
Top Similar × *Post*	0.001	0.007***	0.007***	-0.000
	(1.44)	(3.90)	(3.88)	(-0.77)
Controls	Yes	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes
Observations	39,148	39,148	39,148	39,148
Adj R^2	0.38	0.40	0.59	0.59

This table reports results from firm-level stacked DiD regressions using Sautner et al.'s (2023) measures. In Panel A, the dependent variables are Sautner et al.'s (2023) firm-level exposure measures related to climate change (*CCExposure*), opportunity ($CCExposure^{Opp}$), regulatory ($CCExposure^{Reg}$), and physical ($CCExposure^{Phy}$) shocks. In Panel B, the dependent variables are Sautner et al.'s (2023) firm-level measures of the risk ($CCRisk^{Opp}$), overall sentiment ($CCSentiment^{Opp}$), positive sentiment ($CCSentiment^{Opp,\ Pos}$), and negative sentiment ($CCSentiment^{Opp,\ Neg}$) related to opportunity shocks. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity*. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; *t*-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Appendix A: Variable definitions

Table A.1

Variable definitions.

Variable	Definitions	Data source
CS measure	The percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description.	10-K filings
$\Delta CAPEX/Sales_{i,t+1}$	The change in capital expenditures scaled by sales ($capx/sale$) in years $t+1$ to t.	Compustat
$\Delta R\&D/Sales_{i,t+1}$	The change in research and development expenses scaled by sales ($xrd/sale$) in years $t+1$ to t.	Compustat
$\Delta Div.\ payout/Assets_{i,t+1}$	The change in total dividend payout scaled by total assets (dvt/at) in years $t+1$ to t.	Compustat
VC investment	A dummy variable equal to one if the VC investor finances a climate-tech startup in a given year, and zero otherwise.	PitchBook
Number of deals	The total number of climate-tech startups that the VC investor finances in a given year.	PitchBook
Deal size	The total size of climate-tech startup deals that the VC investor finances in a given year.	PitchBook
VC financed	A dummy variable equal to one if the startup receives VC financing in a given year, and zero otherwise.	PitchBook
CCExposure	Sautner et al.'s (2023) firm-level climate change exposure measure.	Sautner et al. (2023)
$CCExposure^{Opp}$	Sautner et al.'s (2023) firm-level opportunity exposure measure.	Sautner et al. (2023)
$CCExposure^{Reg}$	Sautner et al.'s (2023) firm-level regulatory exposure measure.	Sautner et al. (2023)
$CCExposure^{Phy}$	Sautner et al.'s (2023) firm-level physical exposure measure.	Sautner et al. (2023)
$CCRisk^{Opp}$	Sautner et al.'s (2023) firm-level measure of the risk related to opportunity shocks.	Sautner et al. (2023)
$CCSentiment^{Opp}$	Sautner et al.'s (2023) firm-level measure of the overall sentiment related to opportunity shocks.	Sautner et al. (2023)
$CCSentiment^{Opp,\ Pos}$	Sautner et al.'s (2023) firm-level measure of the positive sentiment related to opportunity shocks.	Sautner et al. (2023)
$CCSentiment^{Opp,\ Neg}$	Sautner et al.'s (2023) firm-level measure of the negative sentiment related to opportunity shocks.	Sautner et al. (2023)
$Top\ Similar_{i,c}$	A dummy variable equal to one if the incumbent firm is treated, and zero otherwise. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. An incumbent firm i is treated in a given VC financing round c if it is identified as a similar firm in that round and not in any other rounds in the preceding four years. An incumbent firm is a control in a given VC financing round if it is in the bottom 20 percentile of similarity scores and is not identified as a similar firm for any other rounds in the entire event window. Control firms are matched to the same 6-digit GICS industry code as the treated firms.	PitchBook; 10-K filings
$Top\ Similar_{i,t-4:t}$	A dummy variable equal to one if a firm i is identified as a similar firm of at least one startup during VC financing rounds in the current year t or in the previous four years ($t-4$ to $t-1$), and zero otherwise.	PitchBook; 10-K filings
Deal/Premoney valuation	The ratio of the total amount of capital invested in the startup in the round to the pre-money valuation of the startup.	PitchBook
ln(Post valuation)	The natural logarithm of the post valuation of the startup.	PitchBook
Generating revenue	A dummy variable equal to one if the startup's business status is classified as "Generating Revenue" or "Profitable", and zero otherwise.	PitchBook
High CoC VC	A dummy variable equal to one if the average cash-on-cash multiple in the four years before the deal date across all VC investors in the syndicate is above the median and at least one VC investor in the syndicate has investment funds where "Climate Tech" or "CleanTech" is a preferred investment vertical according to Pitchbook, and zero otherwise.	PitchBook
Impact investor	A dummy variable equal to one if at least one of the investors participating in the VC financing round is classified as an impact investor by PitchBook, and zero otherwise.	PitchBook
New investors	The number of new investors that participated in the VC financing round for the startup as a proportion of the total number of investors in the round. A new investor is someone who invests in a startup for the first time and has not participated in any prior round of financing for the same startup. We exclude seed rounds when calculating this variable.	PitchBook
Media	The natural logarithm of one plus the total number of news articles featuring the startup from four years prior to the event date to 30 days before the event date.	PitchBook

Table A.1 continued

Variable	Definitions	Data source
Same industry	A dummy variable equal to one if the incumbent firm and the startup share the same 4-digit GICS industry group code, and zero otherwise. We manually match each startup's PitchBook industry code to at least one 4-digit GICS industry group based on the descriptions provided by the respective industry taxonomy.	PitchBook
Existing CS measure	A dummy variable equal to one if the incumbent firm has at least one non-zero value of *CS measure* in the four years leading up to the deal date of the VC financing round, and zero otherwise.	10-K filings
VC state tax	The maximum state-level long-term capital gains tax rate in the headquarter state of the VC investor. This variable is normalized so that a unit change corresponds to a one standard deviation change.	NBER TAXSIM
Firm age	The logarithm of one plus the firm's age, defined as the time between year t and the year in which the firm is first recorded in the CRSP stock database.	CRSP
Firm size	The logarithm of one plus the book value of assets (at).	Compustat
Book-to-market	The logarithm of one plus the book-to-market ratio ($at/(at - ceq + prcc\_f \times csho)$).	Compustat
ROA	Net income divided by total assets (ni/at).	Compustat
Leverage	Total liabilities divided by total assets ($(dltt + dlc)/at$).	Compustat
Sales growth	Current fiscal year sales divided by previous fiscal year sales minus one ($sale_t/sale_{t-1} - 1$).	Compustat
Cash	Cash divided by total assets (che/at).	Compustat
Momentum	Cumulative 12-month return of a stock, excluding the immediate past month.	CRSP
Stock return	The annual stock return of the firm.	CRSP
Stock volatility	The standard deviation of stock returns over the past 12 months.	CRSP
Market share	A given firm's sales divided by the total sales of all listed firms in the same SIC2 industry.	Compustat
Fluidity	The product market fluidity measure constructed by Hoberg et al. (2014). A higher value is associated with a more significant competitive threat for the firm.	Hoberg et al. (2014)

Appendix B: Example of similar and non-similar incumbent firms

In this section, we provide an example of the algorithm used to identify similar and non-similar incumbent firms in a given VC financing cohort. Consider the case of the climate-tech startup NanoCoolers, Inc. Below is the full text of its business description provided by PitchBook:

"Provider of cooling solutions utilizing thermoelectrics. The company develops thermal management cooling solutions that can be applied to a wide variety of areas such as computing, communications, biomedical systems, climate control and refrigeration."

Using the procedure in Section 2.2, the extracted set of keywords from the above business description is: {*thermoelectrics, refrigeration, communications, systems, control, computing, provider, solutions, climate, cooling, thermal, develops, management, variety, areas*}.

On 24 July, 2006, NanoCoolers, Inc received a Series A VC financing round. An example of a similar incumbent firm categorized as a treated observation is Lennox International Inc, as it has a similarity score ranking within the top one percentile among all incumbent firms that filed a 10-K at the fiscal year-end before July 24, 2006. The set of keywords that this incumbent shares with NanoCoolers, Inc is {*systems, management, climate, communications, provider, variety, refrigeration, cooling, solutions, control, areas*}. As expected, this incumbent firm shares a substantial portion of keywords with the startup, given the high similarity score. This similarity can be attributed to the fact that Lennox International Inc is a provider of innovative climate control solutions for heating, ventilation, air conditioning, and refrigeration markets.

In the same cohort, examples of matched control incumbent firms include A. O. Smith Corporation and AAON Inc. Both of these companies are identified as non-similar firms, meaning their similarity scores fall below the 20th percentile, and they share the same 6-digit GICS industry code (201020 "Building Products") with Lennox International Inc. The set of shared keywords of A. O. Smith Corporation and AAON Inc with NanoCoolers, Inc is {*systems, variety, refrigeration*} and {*systems, management, cooling, control*}, respectively. As expected, the keywords of these control incumbent firms have limited overlap with those of the startup, aligning with their non-similar status. However, these two control incumbents operate in the same industry as the treated firm. For example, A. O. Smith Corporation is a manufacturer of residential and commercial water heaters and boilers. Similarly, AAON Inc manufactures heating, ventilation, and air conditioning equipment for commercial and industrial indoor environments.

Another example of a treated similar incumbent firm in the same cohort but in a different 6-digit GICS industry code (251010 "Automobile Components") is BorgWarner Inc. The set of shared keywords with NanoCoolers, Inc is {*systems, management, thermal, develops, communications, provider, variety, cooling, solutions, control, areas*}. This significant overlap is because BorgWarner Inc is a provider of innovative and sustainable mobility solutions in the automotive industry. An example of a matched control incumbent firm in the same GICS industry is American Axle & Manufacturing, Inc. As a non-similar incumbent, this firm has a smaller overlap set of keywords with the startup: {*systems, management, areas*}. However, it operates in the same industry as the treated firm as it is a manufacturer of automobile driveline and drivetrain components and systems.

Internet Appendix

Table IA.1
Strategic investors.

Dep. variable: *CS measure*	(1)	(2)
Top Similar × *Post*	0.267***	0.272***
	(4.71)	(4.71)
Top Similar × *Post* × *Breakthrough Energy*	-0.323	
	(-1.28)	
Top Similar × *Post* × *CVC*		-0.045
		(-0.34)
Controls	Yes	Yes
Firm × Cohort F.E.	Yes	Yes
Year × Cohort F.E.	Yes	Yes
Observations	59,898	59,898
Adj R^2	0.90	0.90

This table reports results from firm-level stacked DiD regressions controlling for other strategic investors. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. *Breakthrough Energy* is a dummy variable equal to one if Breakthrough Energy is one of the investors participating in the VC financing round, and zero otherwise. *CVC* is a dummy variable equal to one if at least one of the investors participating in the VC financing round is classified as a corporate venture capitalist by PitchBook, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity.* For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table IA.2
Early stage climate-tech startup deals.

	Seed	Series A	Seed & Series A
Dep. variable: *CS measure*	(1)	(2)	(3)
Top Similar × *Post*	0.314***	0.174**	0.271***
	(3.79)	(2.34)	(3.98)
Controls	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes
Observations	25,457	13,073	38,530
Adj R^2	0.88	0.92	0.89

This table reports results from firm-level stacked DiD regressions using early stage funding rounds. We focus on an event window of four years before to four years after VC financing rounds. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. An incumbent firm is treated in a given VC financing round if it is identified as a similar firm in that round and not in any other rounds in the preceding four years. If a treated firm is classified as a similar firm in multiple startup deals that occur in the same event year, we assign it to the deal where it has the highest similarity score. An incumbent firm is a control in a given VC financing round if it is in the bottom 20 percentile of similarity scores and is not identified as a similar firm for any other rounds in the entire event window. Control firms are matched to the same 6-digit GICS industry code as the treated firms. We focus on the subsample consisting of Seed rounds in column (1), Series A rounds in column (2), and Seed and Series A rounds in column (3). The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity*. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; *t*-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table IA.3
Alternative stacked DiD specifications.

	Non-zero CS measure		Random match		Control sample cutoffs		
					10%ile	30%ile	40%ile
Dep. variable: *CS measure*	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Top Similar × *Post*	0.156^{**}	0.272^{***}	0.221^{***}	0.260^{***}	0.255^{***}	0.237^{***}	0.233^{***}
	(2.24)	(3.90)	(4.26)	(4.70)	(3.69)	(4.46)	(4.51)
Controls	No	Yes	No	Yes	Yes	Yes	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,343	24,122	23,971	18,919	36,786	79,641	97,301
Adj R^2	0.89	0.90	0.93	0.92	0.92	0.90	0.90

This table reports results from alternative firm-level stacked DiD specifications. In columns (1) and (2), we restrict the sample of incumbent firms to those that have at least one non-zero value of *CS measure* during the event window. In columns (3) and (4), for each treated firm in a given VC financing round, we randomly match it to a firm sharing the same 6-digit GICS industry code but not in the top one percentile of similarity scores to serve as a control. The matched control firm cannot be identified as a similar firm for any other rounds in the entire event window. In columns (5) to (7), the control sample follows the same definition as in the baseline specification but we vary the percentile cutoff that defines this control sample. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity.* For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table IA.4
Differences between firm characteristics using propensity score matching.

	Treatment ($N = 2,974$)	Control ($N = 2,974$)	Difference	
Variables	Mean	Mean	Estimate	p-value
Firm age	2.748	2.746	0.002	0.931
Firm size	6.395	6.501	-0.106	0.227
Book-to-market	0.471	0.464	0.007	0.366
ROA	0.042	0.046	-0.004	0.708
Leverage	0.196	0.198	-0.003	0.751
Sales growth	0.915	4.320	-3.404	0.366
Cash	0.229	0.223	0.006	0.577
Momentum	1.091	1.073	0.018	0.358
Stock return	0.121	0.101	0.020	0.427
Stock volatility	0.137	0.135	0.002	0.531
Market share	0.019	0.020	-0.001	0.745
Fluidity	6.753	6.640	0.113	0.449

This table presents the mean firm characteristics across two subsamples based on propensity score matching. We use one-to-one nearest neighbor propensity score matching without replacement (Roberts & Whited, 2013). We test for differences in the means between the two subsamples and provide the p-values. Standard errors are clustered at the firm-level. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table IA.5
Propensity score matching and weighting models.

	PSM (matched sample)		PSM (WLS)	
Dep. variable: *CS measure*	(1)	(2)	(3)	(4)
Top Similar × *Post*	0.176***	0.186***	0.247***	0.275***
	(3.04)	(2.93)	(5.00)	(5.31)
Controls	No	Yes	No	Yes
Firm × Cohort F.E.	Yes	Yes	Yes	Yes
Year × Cohort F.E.	Yes	Yes	Yes	Yes
Observations	11,993	11,280	60,177	57,321
Adj R^2	0.88	0.88	0.93	0.93

This table reports results from firm-level stacked DiD regressions using propensity score matching and weighting techniques. In columns (1) and (2), for each treated firm in a given VC financing round, we match it to a firm sharing the same 6-digit GICS industry code with the closest propensity score (without replacement) but not in the top one percentile of similarity scores to serve as a control. The matched control firm cannot be identified as a similar firm for any other rounds in the entire event window. In columns (3) and (4), the control sample follows the same definition as in the baseline specification but we use weighted least squares regression with propensity score-derived weights, as in Caliendo and Kopeinig (2008). Treated observations receive a weight of $1/\widehat{p}$, while those in the control group receive a weight of $1/(1-\widehat{p})$, where $\widehat{p}$ denotes the estimated propensity score. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Top Similar* is a dummy variable equal to one if the firm is treated, and zero otherwise. *Post* is a dummy variable equal to one for the event year and subsequent four years, and zero otherwise. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity*. For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Table IA.6
Staggered DiD estimates with heterogeneous treatment effects.

Dep. variable: *CS measure*	(1)	(2)
$Top\ Similar_{i,t-4:t}$	0.133**	0.138**
	(2.40)	(2.35)
Pretrend(-2)	-0.020	-0.017
	(-0.62)	(-0.52)
Pretrend(-3)	-0.044	-0.038
	(-0.99)	(-0.86)
Pretrend(-4)	-0.053	-0.041
	(-1.42)	(-1.13)
Controls	No	Yes
Firm F.E.	Yes	Yes
Year F.E.	Yes	Yes
Observations	14,616	12,523
p-value: All pre-trends are zero	0.278	0.716

This table reports the results using the DiD estimator developed by de Chaisemartin and D'Haultfoeuille (2020), which addresses the issues of treatment effect heterogeneity in staggered DiD regressions. The unit of observation is an incumbent firm–year. Within each VC financing round of a given startup, the set of similar incumbent firms are defined as those in the top one percentile of similarity scores with respect to the startup. $Top\ Similar_{i,t-4:t}$ is a dummy variable equal to one if the firm is identified as a similar firm of at least one startup during VC financing rounds in the current year or in the previous four years, and zero otherwise. The dependent variable, *CS measure*, is the percentage of sentences identified as climate solutions to the total number of sentences in a firm's 10-K Item 1 Business Description. *Pretrend(-k)* is the placebo estimator of de Chaisemartin and D'Haultfoeuille (2020) that estimates the pretrends k years relative to the event year. The omitted category is $k = -1$. We also provide the p-value of the joint test that all pre-trend estimators are equal to zero. Control variables include *Firm age*, *Firm size*, *Book-to-market*, *ROA*, *Leverage*, *Sales growth*, *Cash*, *Momentum*, *Stock return*, *Stock volatility*, *Market share*, and *Fluidity.* For all specifications, standard errors are robust to heteroskedasticity and clustered at the firm-level; t-statistics are reported in the parenthesis. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variable definitions are presented in Table A.1 in Appendix A.

Extract from Lu, Serafeim, et al. (2024) on the *CS measure* creation

Supplementary Note 1: Climate Solutions GPT Model

Data and Sample

Our primary data source is the SEC's EDGAR database, offering public access to 10K filings. A 10K filing is an annual report filed by publicly traded companies in the United States. As a regulatory document, filed with the Securities and Exchange Commission (SEC), companies are required to present factual information, which makes the report more reliable than other sources like sustainability reports and earnings conference calls. The report contains detailed information about a company's overall financial health, business practices, and strategy. Climate solutions are related to the product offering of companies therefore, for our analysis, we specifically targeted the business descriptions found in Part I, Item 1 (Business) of these filings.

Our sample starts with the universe of firms that report SEC 10-K filing in the EDGAR database from fiscal year 2005 to 2022. Our sample period begins in 2005 when the structure of 10-K is more stable. Starting 2005, the Securities and Exchange Commission (SEC) requires firms to disclose the most significant risks in Item 1A (Securities Offering Reform, Item 503(c) of Regulation S-K).

To ensure consistent firm identifiers over time, we use the WRDS-CIK linking tables to map the CIK in 10-K filings to GVKEY in Compustat (Hoberg & Phillips, 2016). This linking table allows us to match firms in Compustat to its historical CIK that could be different from the latest CIK due to firm name and structure changes (e.g., merger and acquisition, spin-offs, and bankruptcies). For example, General Motors filed for bankruptcy in 2009 and received a new CIK following that year. We are able to assign both CIK before and after the bankruptcy to the same GVKEY. We keep firm-year observations that are matched to Compustat as the majority of the firms not matched are funds, which we exclude together with financial institutions since we focus on climate solution products and services, but not the financing of them. Supplementary Table A1 shows the sample composition, where 37% of observations are excluded as a result of this requirement. We then use the Extractor API (Python) from the SEC API to retrieve the raw text of the Item 1 business description section of the 10-K filings. This process results in the loss of around 1% of observations where the API was not able to identify Item 1 or that the identified Item 1 contains fewer than 100 words.

We focus on industries that are pivotal to climate solutions, where our LLM is likely more accurate in identifying climate solutions. Based on reviewing Project Drawdown, we keep 13 (out of 25) GICS industry groups that are central to climate solutions: Energy, Materials, Capital Goods, Transportation, Automobiles & Components, Consumer Durables & Apparel, Food Beverage & Tobacco, Household & Personal Products, Technology Hardware & Equipment, Semiconductors & Semiconductor Equipment, Utilities, Equity Real Estate Investment Trusts (REITs), Real Estate Management & Development. This restriction reduces the sample by 35%. This process results in a final sample of 39,712 observations for 4,485 firms for fiscal years 2005 to 2022.

Climate Solutions Identification

The basis of our metric is a sentence-level binary classifier, designed to detect the presence of climate solutions within the text. This model was specifically developed for sentence-level classification (climate solution or not) due to two primary considerations. First, a sentence, as the fundamental unit of text, presents a clear and concise element for labelers to assess with high accuracy. Second, this method guarantees the precise extraction and identification of text segments specifically relevant to climate solutions.

Supplementary Table 1A: Sample Composition

Sample Composition		
	Sample Size	Ratio
Total 10K from edgar 2005-2022	146,718	
Firms not matched to Compustat	(53,894)	37%
Firms unable to extract item 1	(1,575)	1%
Firms not in relevant industries	(51,537)	35%
Final sample	39,712	

This table shows the sample composition.

Defining Climate Solutions

We define climate solutions as products and services that develop or deploy new technologies in a transition to a low-carbon economy. We identify climate solution technologies based on guidance from the Drawdown Project. The Drawdown Project contains a list of technologies that can reduce greenhouse gases in the atmosphere, and are compiled by a network of scientists and researchers.

While the Drawdown Project provides guidance on what decarbonization technology is considered a climate solution, when we label sentences, we need to decide for which firms the climate solution is a relevant product or service. Consider the following example with three companies involved in the climate solution technology of sustainable aviation fuel (SAF): an energy producer provides SAF to airlines to reduce its emissions and the airline sells flight tickets with lower carbon footprint to a consulting firm. We consider SAF a relevant climate solution for the energy producer since it is the developer of the technology. We also consider SAF a relevant climate solution for the airline since it deploys the technology. However, we do not consider SAF a relevant climate solution for the consulting firm since it engages in business as usual and neither develops nor deploys the climate solution technology.

Creating the training dataset

In the full dataset of almost nine million sentences from 10-K Item 1, only some of them pertain to climate solutions. Therefore, it is crucial to focus on the most representative sentences for efficient training of the model. We select sentences as our training dataset in two steps. In the first step, we select a sample of 100 sentences from each of the 13 industry groups based on sentences most confusing to the model using a one-shot BART model from Setfit. In the one-shot BART model, we predict whether a sentence is a climate solution sentence based on its alignment with Project Drawdown's Solutions Library. By using a BART model instead of randomly selecting sentences, we also ensure a better balance between positive and negative sentences. These chosen sentences go through a labeling process, which we describe in more detail in Supplementary Note 2.

In the second step, we conduct an iterative process to add sentences to the training set through an active learning approach. Active learning is a machine learning technique where the model identifies and selects specific data points for which it requires additional information (labels or annotations) to improve its performance. The technique often involves selecting data points where the model is uncertain. Thus, we identify common types of sentences that our model struggles to interpret or predicts as climate solutions (e.g., sentences considering climate regulations), and we include additional sentences on these confusing areas to further enhance the model. The objective of active learning is to select the data points from which the model learns better, aiming to improve learning efficiency and performance with less labeled data. This approach is particularly useful in scenarios where labeling data is expensive or time-consuming. By focusing on instances where the model's prediction is uncertain, active learning seeks to minimize the amount of required training data, thereby reducing costs and improving the model's accuracy and generalization capabilities.

We use a pre-trained ClimateBERT machine learning model as the base model for the active learning processes (Webersinke et al., 2022). A BERT model has the ability to capture rich contextual information, thus identifying and understanding ambiguous or uncertain cases. This capability enhances the effectiveness of the active learning process by ensuring that the most informative and challenging examples are selected for labeling. The ClimateBERT model's relatively compact size (in its number of weights/parameters) offers the advantage of requiring minimal computational power, enabling comparably quick fine-tuning. To mitigate the drawback of a smaller size model and less context encoded in its weights, the authors of ClimateBERT pre-trained it further on over 2 million paragraphs of climate-related texts to better respond to the domain-specific queries. Like any other binary classification model, ClimateBERT returns a logit, which can be transformed back to probabilities using a logistic function. Based on this output, we conduct the following iterative process:

1. Fine-tuned the model with the data.
2. Choose a decision boundary, that guarantees the highest F1 score.
3. Carefully examine the sentences whose predictions are close to the decision boundary.
4. Use these to guide the addition of new sentences into the dataset.

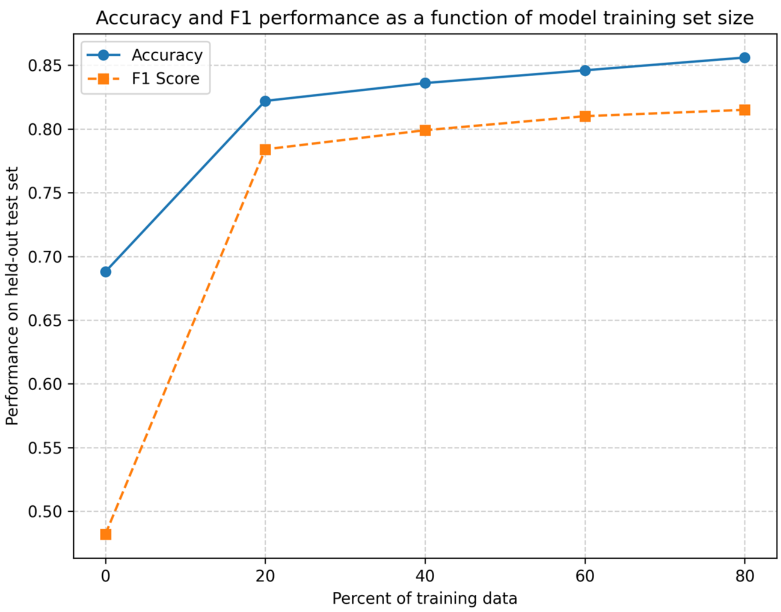
We underwent 8 rounds of active learning and generating training sets, as listed below. For each round, we identify the type of sentences causing confusion to the model and add around 200 sentences to the training set.

1. Sentences that contain "battery" or "electric" but are not related to climate solutions, such as those containing electric toothbrushes.
2. Sentences that describe climate policies or regulations faced by the firm, which does not mean the firm has products or services on climate solutions.
3. Sentences associated with buying carbon credits (e.g., renewable energy credits), but not the creation of carbon credits.
4. Sentences in the building/construction industry that likely needed more examples to properly inform the classifier's decision boundary, specifically when it relates to green buildings and LEED certifications.
5. Sentences containing ethanol, as the model initially does not consider most mentions of ethanol production as climate solution.
6. Sentences where the prefix 'bio' is present, where the model initially classifies as climate solutions but many are not, such as BiOmega-3.
7. Sentences containing generic agricultural products are sometimes misclassified as climate solutions, whereas sentences related to nutrient management and plant-based protein are climate solutions.
8. Sentences containing supporting products to other climate solutions are sometimes not classified as climate solutions. For example, products that enable existing cars to transition to a less carbon-intensive fuel.

This process results in a final training set of 3,508 sentences. The training set statistics are presented in Supplementary Table A2. The size of our dataset is benchmarked to Stammbach et al., 2023, where they annotated 3000 sentences to fine-tune transformer models for climate claim detection (Stammbach, Webersinke, Bingler, Kraus, & Leippold, 2023). Additionally, we evaluate the sufficiency of our training set size by examining how model performance changes as we increase the size of the training dataset. Specifically, we keep a held-out dataset using 20% of the training set, and examine the model performance on this held-out set when we train a GPT-3.5-turbo-1106 model using 0%, 20%, 40%, 60%, and 80% of the training set. Figure A2 shows the largest increase in model performance when the model is fine-tuned with 20% of the training set, compared to the non-fine-tuned model when 0% training data is provided. This increase reflects the value of fine-tuning the GPT model for the specific task of identifying climate solutions sentences. As the proportion of training set increases from 20% to 80%, we do not observe large improvements in model performance, which provides comfort that our training set is sufficient and that we do not anticipate large improvements in model performance if we were to annotate additional sentences.

Supplementary Table A2: Composition of the training data

Industry Name	Count in Training Set	Number of Positives	% of Positives	Count Overall	% of the Training Set	% of Overall Set
Automobiles and Components	291	168	0.577	180,942	8.640	1.984
Capital Goods	405	139	0.343	1,238,834	12.025	13.588
Consumer Durables and Apparel	178	45	0.253	494,283	5.285	5.421
Energy	181	74	0.409	1,769,351	5.374	19.406
Equity Real Estate Investment Trusts	188	57	0.303	492,869	5.582	5.406
Food, Beverage and Tobacco	451	163	0.361	412,178	13.391	4.521
Household and Personal Products	146	18	0.123	220,420	4.335	2.418
Materials	301	71	0.236	896,885	8.937	9.837
Real Estate Management and Development	134	40	0.299	492,869	3.979	5.406
Semiconductors and Semiconductor Equipment	178	69	0.388	460,569	5.285	5.052
Technology Hardware and Equipment	158	33	0.209	917,277	4.691	10.061
Transportation	184	46	0.250	313,834	5.463	3.442
Utilities	573	331	0.578	1,227,056	17.013	13.458



Supplementary Figure A2: Model Performance relative to Training Size

Training Methodology and Model Selection

We use the labeled training set to fine-tune a GPT-3.5-turbo-1106 specialized at labeling climate solutions sentences. Fine-tuning is the process of further training a pre-trained GPT model on a specific data and involves adjusting the model's weights to better capture the language and concepts related to climate solutions. GPT algorithm is based on a neural network architecture that depends on weights, which are the parameters that are learned during training. The weights determine the strength of connections between neurons in different layers of the model. Adjusting these weights changes the way the model processes input data and generates output. Fine-tuning adjusts the model's weights so it can better understand and generate climate-specific terms and phrases, such as "renewable energy," "plant-based protein," and "cogeneration." Through this process, the model learns the contextual usage of these terms within climate-related discussions, improving its ability to generate relevant and coherent text specific to climate solutions. The fine-tuning hyperparameters for our GPT-based model are based on recommended defaults, with epochs set to 3, batch size to 7, and a learning rate multiplier of 2.

We employ 5-fold cross-validation to assess our model, optimizing the use of our labeled dataset. This method ensures comprehensive evaluation by partitioning the dataset into five subsets, where each subset serves as a test set while the remaining are used for training, iteratively. For each fold, we designate 20% of the labeled dataset as a holdout set for testing, while the remaining 80% is used to fine-tune a GPT-3.5-turbo-1106 model. The trained model is then evaluated on the held-out 20%, and this process is repeated across all five folds.

The model demonstrates an overall accuracy of 84.09%, with a standard deviation of 1.93% between folds, indicating consistency in performance across different subsets. Moreover, we report an F1 score of 79% with a standard deviation of 2% between the folds. The F1 score, being the harmonic mean of precision and recall, provides a balanced measure of the model's accuracy, particularly valuable in the context of binary classification. It is especially pertinent for evaluating performance in imbalanced datasets, where traditional accuracy metrics may not fully capture the effectiveness of the model in distinguishing between classes.

We display the GPT prompt below and the detailed model performance by industry in Supplementary Table A3.

Listing 1: GPT finetuning prompt

```
system_message = ''You are a chatbot with expertise in
   environmental regulations and climate change mitigation
   strategies. Your function is to meticulously analyze
   sections of regulatory documents, 10k filings, to identify
   the presence of proposed climate solutions. Based on the
   guidelines, assess whether the company is implementing
   specific technologies or practices contributing to a low-
   carbon economy. Look for whether there is a clear
   indication of the company's investment or future investment
    in climate solutions or the sentence implies a reduction
   in carbon emissions through the company's products or
   services. Generic, vague, or general statements about
   climate change should classified as no.''
```

Supplementary Table A3: Model Evaluation by Industry

	GPT 3.5 FT		ClimateBERT FT	
Index	**F1 Score**	**Accuracy (%)**	**F1 Score**	**Accuracy (%)**
Automobiles and Components	0.845	81.787	0.836	80.756
Utilities	0.839	81.501	0.839	80.279
Real Estate Management and Development	0.835	90.299	0.706	81.343
Energy	0.833	86.740	0.833	85.635
Food, Beverage and Tobacco	0.793	83.814	0.799	84.257
Capital Goods	0.789	84.938	0.772	82.963
Consumer Durables and Apparel	0.776	87.640	0.729	85.393
Technology Hardware and Equipment	0.727	88.608	0.646	85.443
Semiconductors and Semiconductor Equipment	0.724	76.405	0.723	75.843
Transportation	0.722	85.326	0.667	82.065
Materials	0.699	85.382	0.649	82.060
Equity Real Estate Investment Trusts	0.694	80.319	0.625	74.468
Household and Personal Products	0.636	89.041	0.528	82.877
Overall	**0.795**	**84.090**	**0.776**	**81.984**

In the process of selecting an appropriate LLM, we considered several aspects:

- **Cost:** The financial implications of model utilization vary significantly depending on the deployment strategy. For models operated on private infrastructure, the primary cost consideration involves the expenses associated with cloud services. Alternatively, when employing a proprietary model accessible via API, the cost per token becomes a pivotal factor. Opting for the latter, our strategy focused on crafting concise prompts to minimize expenses without compromising the model's effectiveness.

- **Latency:** The response time of models can range widely, influenced by factors such as model size, architectural complexity, and the computational power of the hosting environment. This variance is a critical consideration, especially in scenarios requiring rapid iterative testing and feedback. Although larger, more computationally intensive models may offer superior performance, selecting a model that balances response time and computational demands was essential for our workflow. In our approach, we utilized both, the efficiency of climateBERT to rapidly iterate over training examples and then the large context of a GPT model for the final classification.

- **Performance for Specific Tasks:** The adaptability of modern LLMs to a broad spectrum of tasks is remarkable, often eliminating the need for fine-tuning or complex prompting strategies for general applications. However, specialized tasks may necessitate tailored adjustments or fine-tuning to achieve optimal results. The trade-off between using generalized versus specialized language models for niche domains has been explored in research, such as in medicine (Nori et al., 2023) and finance (Li et al., 2023).

Given these considerations, our choice of the final model was informed by a holistic assessment of primarily task-specific performance, cost, and latency. Despite ClimateBERT's suitability for our initial training needs, the superior performance, expansive knowledge, and versatility of GPT were sufficient reasons to be our choice of model for the final classification phase. Looking at the overall accuracy, the fine-tuned GPT-3.5 is at 84.09%, which is higher than that of the fine-tuned ClimateBERT's accuracy of 81.98% as seen in Supplementary Table 1C. We select the ClimateBERT model with the highest F1 score by conducting a grid search over key hyperparameters, including learning rates (5e-05, 2e-05, 1e-05, 5e-06), epsilons (1e-08, 1e-07), and dropout probabilities (0.1, 0.2, 0.3). The optimal model for which our accuracy and F1 scores are based on has a learning rate of 5e-05, epsilon of 1e-08, and dropout of 0.1.

In untabulated analysis, we also explore three alternative models, DistilRoBERTa, RoBERTa, and DeBERTa (He, Liu, Gao, & Chen, 2020; Liu et al., 2019; Sanh, Debut, Chaumond, & Wolf, 2019), with the same parameters as the ClimateBERT model. Across these models, the F1 and accuracy rates are below the ClimateBERT model, which has a performance below the fine-tuned GPT we use.

In particular, the fine-tuned GPT-3.5 outperforms the fine-tuned ClimateBERT significantly in correctly identifying climate solutions sentences in industries with fewer climate solutions, such as in Equity Real Estate Investment Trusts and Household and Personal Products. In summary, by fine-tuning GPT-3.5-turbo-1106 with a targeted training set, we achieve a balance between cost efficiency and performance.

A challenge with utilizing GPT-3.5 is its non-deterministic behavior. Non-deterministic behavior in GPT refers to the variability in its outputs even when given the same input multiple times. This behavior arises from several factors inherent to the design and operation of the model. One key factor is temperature, which controls the randomness of predictions. A higher temperature value (e.g., 1.0) produces more random outputs, while a lower temperature (e.g., 0.1) makes the output more deterministic and focused on high-probability tokens (refer to the words or sub-words that the model predicts are most likely to come next in a given sequence). Therefore, to reduce variability in predictions, we set the temperature hyper-parameter to 0.1. Additionally, to examine the potential variability in this non-deterministic behavior, we randomly selected 1,000 sentences from outside the training set and apply the fine-tuned GPT model five times. The maximum discrepancy observed between any two columns was 1 row.

Supplementary Note 2: Climate Solutions Labeling

To train our GPT climate solutions model, we label 3,508 sentences as either climate solution sentences or not. For our annotation procedure, we implement the following general rules referencing Webersinke et al. (2022). The annotators have to determine whether a sentence is related to climate solutions. Annotators are asked to apply common sense, e.g., when a given sentence might not provide all the context, but the context might seem obvious. Moreover, annotators are informed that each annotation should be a 0-1 decision. Hence, if an annotator is 70% certain, it is rounded up to 100%. Two researchers annotate the same tasks to obtain some measure of dispersion. In case of a close verdict or a tie between the annotators, the authors of this paper discuss the sentence in depth before reaching an agreement. Out of 3,508 sentences, annotators agreed on 2,905, while the remaining sentences had disagreements. To assess the degree of annotator agreement, we calculate Cohen's Kappa, which is 0.6653 with a 95% confidence interval of 0.64 to 0.6907. This indicates a substantial level of agreement in the labeling process.

We define climate solutions as products and services that develop or deploy new technologies in a transition to a low-carbon economy. As a general rule, we determine that just discussing climate change or the environment is not sufficient, the sentence should mention specific climate solutions, such as renewable energy, electrification of transportation and processes, battery technology, new agricultural practices, or plant-based protein alternatives to meat. When in doubt, we refer to the list of climate solutions technologies listed in Project Drawdown. Below, we provide some examples.

Sentence	Label	Reason
Our industry experience, the performance of our transit buses, and compelling total cost of ownership has helped make us the leader in the U.S. electric transit bus market.	1	The firm is creating electric transit bus, which, as an electric vehicle, is a climate solution.
We believe we have a responsibility and opportunity to play a role in the global economic transition to net zero emissions.	0	This is a generic statement without referencing specific products or investments, as compared to the previous sentence.
Our expanding corporate offices in Los Angeles, California are being designed and developed to qualify for LEED certification.	0	This is about their current operations, and not a product they are developing.
Many of our products meet the requirements for the awarding of LEED credits, and we are continuing to develop new products, systems and services to address market demand for products that enable construction of buildings that require fewer natural resources to build, operate and maintain.	1	This is similar to the last sentence in mentioning the LEED certification, but is used with respect to a product, and therefore qualifies.
The first class of QFs includes energy producers that generate power using renewable energy sources such as wind, solar, geothermal, hydro, biomass or waste fuels.	0	This reads as part of a regulation for Qualifying Facilities, and not a product or any indication of a company's actions.

Additional references

He, P., Liu, X., Gao, J., & Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654.

Li, X., Chan, S., Zhu, X., Pei, Y., Ma, Z., Liu, X., & Shah, S. (2023). Are chatgpt and gpt-4 general-purpose solvers for financial text analytics? a study on several typical tasks.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., . . . Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692, 364.

Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., . . . Horvitz, E. (2023). Can generalist foundation models outcompete special-purpose tuning? Case study in medicine.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Stammbach, D., Webersinke, N., Bingler, J., Kraus, M., & Leippold, M. (2023). Environmental claim detection. Association for Computational Linguistics, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 1051–1066.

Webersinke, N., Kraus, M., Bingler, J. A., & Leippold, M. (2022). ClimateBert: A pretrained language model for climate-related text. Working Paper.