Three Essays in Investments

Jinyoung Kim

A dissertation

submitted to the Faculty of

the Department of Finance

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College
Wallace E. Carroll School of Management
Graduate School

April 2024

Three Essays in Investments

Jinyoung Kim
Department of Finance, Boston College

Dissertation Committee:
Professor David H. Solomon (Chair)
Professor Jeffrey Pontiff
Professor Jonathan Reuter
Professor Ran Duchin

Abstract

My dissertation comprises three essays delving into questions that contemporary investors encounter in the ever-evolving landscape of investments. The first essay examines how the presence of public pension funds as limited partners influences venture capitalists' (VCs) risk-taking behaviors. It notes that investments by public pension funds in the venture capital market have increased over the past two decades, and these funds possess unique objective functions compared to other venture capital investors. Findings suggest that VCs backed by public pensions tend to invest in startups with lower-risk profiles, such as those with technologies related to public companies, numerous patents, and later funding rounds, leading to more frequent and quicker exits but lower returns. To establish causality, I employ an instrumental variable evaluating the likelihood of public pension funding based on the location of funds initiated during a typical fundraising cycle in a venture capital firm. Furthermore, I find that public pensions prefer venture capital firms with a track record of conservatively managing funds, particularly those pensions that have previously engaged with such firms.

The second essay shifts focus to the stock market, documenting higher returns from companies developing new technologies. The advancement of new technologies is pivotal to an economy's potential, yet it carries inherent risks. As per investment theories, investors demand premiums for holding stocks associated with high

uncertainty, prompting questions about whether they are adequately compensated for investing in companies undertaking highly uncertain projects. A novel application of a graph-neural network model identifies new technology patent publications annually, enabling the calculation of firms' exposure to new technologies. With the measure, I find that portfolios with high new-tech exposure outperform those with low exposure, driven by significant risk premiums. This sheds light on the positive correlation between idiosyncratic risk and stock returns, contributing to our understanding of the market's valuation of technological innovation.

The third essay presents a systematic analysis of stock market valuations of Corporate Social Responsibility (CSR) initiatives. The study identifies public demand for CSR as a pivotal factor in enhancing the value of CSR activities. Analyzing market reactions to CSR activities via cumulative abnormal returns, the research finds overall neutral market responses. Nonetheless, it finds that heightened public concern for specific issues can sway market reactions positively. Also, when CSR initiatives employ strategies that extend beyond the capabilities of individuals, the market responses tend to be favorable. The paper further shows that firms strategically increase their CSR activities and choose implementation modes, aiming to enhance their value. To explain why market reactions are, on average, neutral, I further provide evidence suggesting reasons such as virtue signaling, a lack of understanding of the importance of profitability, and other executive motives.

Together, these essays deepen our understanding of investments by exploring how financial market participants, corporate endeavors in technological advancements, and societal expectations for corporate social responsibility influence investor behavior and asset prices.

Acknowledgements

This endeavor would not have been possible without the invaluable support from faculty, staff, and fellow students in the Finance Department at Boston College. I am deeply thankful to my dissertation committee chair, David H. Solomon, whose expertise, compassion, and insightful guidance were pivotal throughout my journey at Boston College. Having him as my advisor was one of the most fortunate occurrences during my time in the program. His intellectual curiosity, openness, and wit significantly impacted my research experience. My deepest gratitude also extends to Jeffrey Pontiff, Jonathan Reuter, and Ran Duchin, especially for their generosity. Engagements with them have been immensely enriching and helpful. Additionally, I wish to express my heartfelt gratitude to the professors who showed warmth and support in various forms. Particularly, I want to thank Thomas Chemmanur for his sporadic encouragement and interactions, which have shown me how even small gestures of support can meaningfully impact someone's resolve to press on. Additionally, I thank my former office mates, Xiang Li and Jiajie Xu, for demonstrating diligence and passion in their research and for their kindness, from which I learned a great deal. Finally, I express my gratitude towards the Department of Finance at the Carroll School of Management for their financial assistance throughout this journey.

On a personal note, I extend my heartfelt thanks to my parents and sister, whose love and support have been instrumental in enabling me to pursue this work. I am also grateful to my friend, Han Bitsongie, who was always there to answer my calls and provide comfort during challenging times. My sincere thanks to them for everything.

Table of Contents

1 The Impact of Public Pension Funds as LPs on VC Investments

Over recent decades, there has been a noticeable increase in investments from public pensions into alternative investments including the venture capital market.[1] While many have examined why public pensions have increased their exposure to alternative investments and the potential ramifications of this shift for them, few have explored the impact on venture capitalists (VCs) who have seen a surge from this unique group of investors.[2]

The dearth of research on the impact of public pension investments on VCs is particularly puzzling for two reasons. First, investors in the venture capital market, often referred to as limited partners (LPs), are more than just capital providers for the current fund of a VC. They represent potential sources of future capital and serve as channels for building and preserving the VC's reputation, which provides VCs with incentives to maintain a good relationship with their investors.

Specifically, investors in the venture capital sector are often tied to long-term, significant commitments. This leads VCs to incur the costs of finding new investors willing to contribute to their funds (Duffie, Gârleanu, and Pedersen (2007)). However, when dealing with familiar investors, the information asymmetry is lower, resulting in decreased search costs. Consequently, these known investors are considered potential contributors to the VCs' follow-on funds. Also, the past and current investors of a VC can act as a significant source of information about the VC for other investors and startups. This becomes particularly important given the opacity of the venture capital market, where information is less accessible. As reputation plays a pivotal role in the success of VCs (Gompers and Lerner (1999), Nahata (2008)), investors as a reputation channel are important for VCs.

Given their substantial asset size, public pensions hold the potential to be repeat investors. Furthermore, they typically have an extensive network of connections which can shape a VC's reputation. As a result, they are likely to be regarded as important

[1] From 2006 to 2016, public retirement systems more than doubled their allocations to alternative investments such as private equity, hedge funds, and venture capital funds, growing from 11% to 26% (Pew Research Center (2018)). Gompers and Lerner (1999) reported that 9% of total VC funding came from public pensions in 1999, utilizing the Venture Economics database and hand-collected information.

[2] Some views on this trend attribute the shift to pressures from underfunded goals (Pennacchi and Rastad (2011), Mohan and Zhang (2014), Lu et al. (2019)), the desire to achieve specific return targets (Andonov, Bauer, and Cremers (2017)), and a changing belief in the risk-reward profiles of alternative investments (Begenau, Liang, and Siriwardane (2023)).

clients by VCs.

Secondly, there is a need for research on public pension investments in the venture capital market as the financial implications of risk-taking by VCs differ for public pension funds compared to other investors. The primary goal of public pension plans is to provide retirement income to their public employees. State and local retirement systems predominantly offer defined benefit schemes. As noted by Brown and Wilcox (2009), the robust contractual and legal protections often render public pension benefits nearly risk-free and these obligations can be calculated using a variety of actuarial methods. This contrasts sharply with the nature of returns in the venture capital market, characterized by significantly skewed returns and substantial fluctuations (Cochrane (2005), Kerr, Nanda, and Rhodes-Kropf (2014)).

Importantly, operating under a defined benefit system provides a unique objective function for public pension funds. Unlike most venture capital investors, who are not subject to additional financial liabilities if their investments lose value, public pensions face heightened financial responsibilities when confronted with significant negative returns. On the contrary, estimated financial obligations cap the advantages gained from extremely high investment returns. As a result, public pensions may have a lower risk appetite than other venture capital investors, owing to their objective function. Hence, the same risk profile of a venture capital fund may provide different utility to public pension funds compared to other investors.

To explore the potential impact of public pension investments on VCs, this study focuses on the VCs' risk-taking behavior in the context of accommodating their investors' preferences. Specifically, I propose that managers with public pension investors form less risky portfolios, which are marked by a higher likelihood of achieving payoffs, albeit without the potential for exceptionally high returns. Essentially, the hypothesis suggests that investors' risk preferences could be a factor that VCs consider when making investment decisions.

Three key pillars support the risk-taking catering hypothesis. Firstly, investors play a pivotal role in the success of venture capital firms as future capital providers and channels for building reputation. This incentivizes VCs to cater to their investors. Secondly, despite public pensions increasingly allocating funds to riskier asset classes such as equity or alternative investments, their risk tolerance remains lower than other VC investors due to defined benefit systems and pension governance. This suggests that catering by VCs likely follows a specific direction for public pension funds. Lastly, VCs possess the ability to modulate the risk levels of their portfolios, given that each investment constitutes a significant share of their total portfolio. This indicates that

VCs have the capacity to align their investment strategies with the preferences of their investors.

The study focuses on VC investment-level risk metrics inspired by trends in exit markets and findings in the literature. Firstly, it assesses the degree to which a startup's business aligns with public companies, noting that acquisitions have become a common yet increasingly less rewarding exit strategy from 2001 to 2022. [3] Secondly, the study evaluates the number of patents held by a startup, viewing this as an indicator of the startup's appeal as an acquisition target and a measure of its ability to protect future revenues. This aspect not only boosts the startup's valuation but also impacts the decision to pursue an IPO.

I formally show that these two metrics are associated with a high likelihood of exits. In the sample of startups funded by U.S. VCs between 2001 and 2017, where the last investment was made before 2018 as of 2023, I find that startups with technology more closely linked to public companies are more likely to exit via acquisition. The number of patents strongly predicts successful exits through both acquisitions and IPOs.

Two other investment-level risk metrics are based on earlier work. I use a dummy variable for early-stage investments and a dummy variable for startups that have received late-round financing in recent years. Existing research suggests that late-stage investments are generally less lucrative but could be appealing due to their higher frequency and swifter timelines for successful exits (Sorensen (2004), Cochrane (2005), Chaplinsky and Gupta-Mukherjee (2016)).

Using these metrics, I investigate how the presence of public pensions influences the risk profile of VC investments. I employ an indicator variable to denote the participation of a public pension investor in a VC fund. In the baseline regression, where an investment-level risk metric is regressed on the indicator variable, the involvement of public pensions is related to lower-risk investments. Specifically, the indicator variable for public pension funds shows a positive association with startups' relatedness to public companies and the number of patents. Additionally, their presence is associated with investments in startups at later stages.

To examine the effect of the presence of public pensions, I need to address endogeneity issues. Notably, public pensions can choose certain types of VCs. To navigate this issue, I adopt an instrumental variable (IV) approach. I create an instrumental variable that makes public pension investments likely unrelated to public pension se-

[3] In my sample, 72.1% of successful exits are acquisitions. This is similar to 65.83% in Barrot (2017).

lection or risk appetite. The variable is an indicator that represents funds that were created as part of VCs' typical fundraising cycle and experienced a high likelihood of inflows from public pension funds due to their location.

The conventional fundraising cycle is defined based on a practice that is discussed by Barber and Yasuda (2017). In practice, VCs initiate the capital acquisition for a new fund between the third and sixth years of the incumbent fund's operational lifespan. This is the period during which managers can showcase their investment skills through the interim performance metrics of the existing funds. Therefore, I posit that VCs are poised to attract new investments from various investors during those times.

The locations where individual VC funds can anticipate inflows from nearby public pension funds can be divided into two elements: the geographical proximity of VC funds to public pensions and the expected size of inflows from these pensions.

Public pensions may prefer investing in geographically closer VC funds to reduce supervisory costs. Lower costs stem from both informational benefits and reduced travel expenses (Coval and Moskowitz (2001), Tian (2011), Ellis, Madureira, and Underwood (2020)). Additionally, public pensions investing in nearby VC funds is consistent with their broader objective of invigorating regional economies, as these VC funds are also more likely to invest in geographically proximate startups. My analysis of interstate investment patterns of public pension funds supports this argument. VC funds located in the five geographically nearest states, including the public pension fund's home state, are more likely to receive an investment from that pension.

Leveraging this concept, I calculate the collective potential funding that public pension funds within a designated state might extend to adjacent VC funds, premised on the notion that capital is uniformly distributed among VC funds in close proximity to these pension funds. For this analysis, "adjacent" or "neighboring" VC funds are defined as those situated within the state or in the five nearest states including the designated state. Initially, I determine the discrepancy between targeted and actual private equity allocations by public pensions in a state and then distribute this difference across the number of adjacent VC funds. If the resultant value ranks in the top 25th percentile of pensions' state-year observations within the dataset, it indicates that, during such state-years, VC funds located in the state and its neighboring states are likely to secure funding from public pensions.

The IV is constructed by intersecting the two components. The IV is an indicator that takes a value of one if a VC fund was established due to the typical VC fundraising cycle, and its creation was in a state experiencing a favorable chance of receiving funding from public pensions as of the capital raising phase.

With the IV, I conduct two-stage least squares estimations with saturated covariates to find a local average treatment effect (Angrist and Imbens (1995), Blandhol et al. (2022)). In the second-stage regression analysis, the data supports the hypothesis that VC managers adopt more conservative portfolio strategies when public pension funds are among their investors.

Building on the observation that VCs with public pension investors tend to choose investments associated with a higher likelihood of exits, though possibly at lower returns, I investigate if this investment strategy does extend to their exit likelihood. I find that VCs with public pensions are more likely to achieve successful exits of their investments, reinforcing the idea that VCs strategically manage their portfolios to increase the chances of payoffs.

Moreover, VCs backed by public pensions are geared towards accelerating the exit process. I introduce a variable that measures the time, in years, it takes for an investment to exit following the establishment of a VC fund, within the fund's lifespan. The results indicate that VCs, when working with public pension investors, are inclined to make investments that result in exits occurring, on average, 1.7 years sooner than those made by other venture capitalists with funds initiated in the same year.

I further carry out additional analysis to deepen the insights gained from the findings. Firstly, I investigate the relationship between the presence of public pension investors in a VC firm and the firm's approach to risk across its entire investment portfolio, including VC firm fixed effects. I find that the presence of public pensions does not significantly alter a firm's overall risk propensity. Only those funds invested in by public pensions exhibit less risky portfolios. This suggests that VCs strategically distribute their desired level of risk across the various funds they manage, rather than entirely shifting the firm's risk management style.

Next, I explore the association between investment-level risk indicators and returns, to verify if less risky investments correlate with lower returns, as hypothesized. I examine round-to-round returns and find that the investment-level risk metrics are related to the returns in a way that is consistent with the risk-reward trade-off.

I further examine how public pension funds make investment decisions in the venture capital market. Given that the fundraising phase for a VC fund occurs before the selection of investments, public pensions are unable to assess the risk profile of the fund at the time of their investment. Consequently, I posit that public pensions deduce the risk management approach of a VC fund through an examination of the firm's historical investment activities. I find that if a VC fund is affiliated with firms having a track record of a conservative investment style, the fund is more likely to have public pensions as investors.

I delve deeper into the findings by dissecting the investment patterns of public pension investors, distinguishing between newcomers to a VC firm and seasoned investors with a history of engagements with the firm's past funds. To this end, I create a dataset of pairs of public pension investment units and VC funds available for investment when these pensions made their decisions. I find that first-time public pension investors in a VC firm do not place significant emphasis on the risk profiles of the firm's previous investments. In contrast, pension funds with prior experience investing in a VC firm tend to reinvest in the firm's future offerings if the firm has demonstrated a preference for less risky investments. The analysis indicates that public pensions appear content with the outcomes produced by a conservatively managed portfolio. Therefore, VCs are effectively achieving the objective of maintaining a good relationship with public pensions through portfolio management.

Next, I investigate whether the financial health of public pension investors explains their selection of VC funds. My findings suggest that the funded status or a gap between the assumed rate of returns and past investment returns is not markedly associated with the VC funds they opt for. However, there is a slight indication that underfunded public pensions appear to favor VC firms with a track record of less risky investments. This inclination might elucidate why underfunded pensions witness smaller returns on their private equity investments (Mittal (2022)). These seemingly modest returns could signify a preference for lower risks.

In summary, the results presented in this paper demonstrate that the presence of public pension investors affects VCs' risk-taking. The finding introduces an additional determinant of VCs' risk management. Previous studies have highlighted that contractual incentives shape managers' approach to risk (Buchner and Wagner (2017)), GPs' capital commitments (Jia and Wang (2017)), market cycles (Nanda and Rhodes-Kropf (2013)), and factors such as the size and age of the venture fund also influence their risk-related decisions (Giot, Hege, and Schwienbacher (2014)). This paper shows the characteristics of investors can influence VCs' investment decisions. This is an interesting finding since it challenges the prevalent view that investors have little impact on how fund managers manage risks. In the VC industry, investors might exert more influence on VCs, given the smaller number of qualified investors and their relatively close interaction with the managers, when compared to other asset management industries. Most importantly, VCs rely on the existing relationship with their investors to secure funding for their follow-on funds and maintain their reputations.

The paper is also related to the literature on how VCs can affect the technological landscape. Several research papers have established that VCs can play a crucial role in fueling innovation (Nanda and Rhodes-Kropf (2013), Pierrakis and Saridakis (2017), González-Uribe (2020)). I show in the study that VCs can influence the direction of technological progress. Specifically, the research brings to light an important implication. If VCs are backed by investors who need prudent risk management, GPs invest in technologies that are already proven to be worth exploring by public companies. On the other hand, if VCs are backed by investors who exhibit a high degree of risk tolerance, they are likely to go for more exploratory innovations. This willingness to embrace risk and uncertainty can lead to groundbreaking and transformative innovation. Thus, the predominant investor type in the VC industry can shape the overall technological trajectory of startups.

The remainder of the paper is structured as follows: Section 1 introduces data used in the study. Section 2 shows the relationship between technologies and patents of startups and their exit likelihood. Section 3 investigates the effect of public pension funds on VC investments. Section 4 provides additional analyses enhancing the understanding of the main results, and Section 5 concludes.

1.1 Data

The primary data used in this study is gathered from VentureXpert, a venture capital database provided by Thomson Economics. I download all venture capital investments (pairs of a startup and a VC fund) from 1980 to 2022 that were made by U.S. VC firms. The database provides information on the stage of each deal. I loosely refer to both seed-stage and early-stage deals as early-stage investments. The data also contains the timing of deals. For each startup, I identify the most recent VC investment date, which will be used to pinpoint startups nearing the formulation of their exit strategy. I enrich the VC investment data by integrating information from multiple sources for my research.

1.1.1 Startups' IPO and M&A information

VentureXpert provides information regarding the public status and company status of startups backed by a VC fund listed within it. However, in some cases, it retains outdated data, and the database does not provide comprehensive information on the exit dates.[4] Therefore, I gather more information from other databases that extensively cover M&As and IPOs. Specifically, I consult the SDC Global New Issues database and the SDC's Mergers and Acquisitions database. I focus on completed IPOs and acquisitions by corporate acquirers, which serve as reference points for determining the startup's exit date. In addition to the information, I augment the data with the IPO date provided by VentureXpert.

To combine VentureXpert data and the two datasets on IPOs and M&A, I employ the tiered, exact name-matching approach explained in the Technical Appendix. The tiered, exact name-matching approach involves finding counterparts in different databases based on exact name matches while gradually allowing for variations in names. This method is used to increase the probability of matches while maintaining the quality of the matches.

After the matching process, I identify the earliest event (whether an acquisition by corporate acquirers or IPO) for a venture as the exit event, using a set of events from SDC combined with IPO dates from VentureXpert. After the step, I find that 18,242 startups exited through acquisitions, and 4,897 exited through IPOs among the 74,158 startups in the sample.

On top of the exit events identified in the step above, I also make use of information supplied by VentureXpert. Two fields in VentureXpert detail an investee company's current status: 'public status' and 'company status.' If a venture is marked as 'Subsidiary' in the public status field and 'Acquisition' in the company status field, I interpret this as an exit through an acquisition, even if I can't find the venture in either of the M&A and IPO databases. Similarly, if a startup is labeled 'Public' in the public status field and 'Went Public' in the company status field, I conclude that the company exited via an IPO. I make these adjustments because even though VentureXpert might provide stale data for these fields, the database is unlikely to present false negatives for exits since the default status for a venture is likely to be an active, private company. For those additional exits, I assume startups exit five years post the date of the most recent investment from a VC fund, based on the last VC deals recorded in VentureXpert.

As a result of the additional step, there are 3,318 additional startups that have exited through acquisitions, and 1,202 more startups that have exited through IPOs. Consequently, acquisition-driven exits constitute 29.1 percent of the sample startups,

[4]For example, as of April 2023, Burlington Industries is listed as an active private company in VentureXpert and is not classified as a subsidiary. However, the firm was acquired by International Textile Group in 2003.

while IPO-driven exits account for 8.2 percent during the sample period from 1980 through 2022. By narrowing down the sample to startups that received their last funding from a VC fund before 2018, providing sufficient time for these startups to mature for potential exits, I discover that 29.7 percent of startups have exited through acquisitions, while 11.5 percent have undergone IPOs.

1.1.2 Patent Information of Startups

This paper explores the technological orientation of startups, specifically in relation to how it connects with technologies developed by public companies. To assess the technological focus of these startups, I draw on the patenting activities of startups and public firms, sourcing this information from Google Patent. In this subsection, I outline the method used to identify patent information associated with startups, while the process for identifying patent data pertaining to public companies is detailed in the Technical Appendix.

The United States Patent and Trademark Office (USPTO) discloses patent application publications (pre-grant publications) and patents as part of their patent granting process. In addition to granted patents, the USPTO publishes patent applications that have not yet been granted as patents. Most patent applications are published 18 months after their filing date, or earlier in certain circumstances. I refer to both types of publications as patent publications.

In order to identify the patent publications of startups, I need two specific dates for each startup: the inception date and the date when the venture either successfully exited or discontinued its operations. These dates enable me to determine all inventions submitted to the USPTO within this timeframe as the patenting activities of the given venture.

For startups that have successfully exited, I utilize the exit date defined in the aforementioned subsection. For companies that have ceased operations, I first identify these startups using the 'public status' and 'company status' fields in VentureXpert. A venture is considered defunct if the 'public status' field indicates 'Defunct', or the 'company status' field displays 'Defunct', 'Bankruptcy - Chapter 11', or 'Bankruptcy - Chapter 7'. I presume that the companies became bankrupt five years after receiving their last investment from a VC fund. The choice of a five-year period is based on the common practice of VC funds selecting startups in the first five years of their lifecycle and writing off their investments within a decade of the fund's establishment (Tian and Wang (2014)). Rather than selecting the longer ten-year timeframe, I choose the more conservative five-year period to align patent data with startups, especially considering the possibility of new startups with identical names filing for patent applications.

For firms that have not exited or become defunct, I presume they are still operational as of the end of 2022. In such cases, I attribute all inventions published prior to December 31, 2022 to these startups. If a startup has a specified founding date, all inventions disclosed after this date are associated with the startup. For startups lacking a specified founding date, I attribute all inventions published post-January 1st, 1980 to the startup.

It is common for the same invention to be published multiple times by the patent office, for instance in the case of updated filings for the same invention. In order to avoid counting the same invention multiple times, I identify a unique invention by its earliest assigned priority number. This approach ensures each invention is counted only once, no matter how many times the patent office has published it, offering a precise timeline of when the unique invention was developed by its inventor. While I continue to use terms like 'patent applications' or 'patent references' in subsequent sections, it is essential to understand that they refer to unique inventions and not repeated publications by the patent office.[5]

The process results in 541,763 unique patent publications and 248,300 unique inventions matched to 19,167 startups.

1.1.3 Identifying Public Pension Funds as Limited Partners

To compile historical records of public pension funds serving as limited partners, I manually gather data on private equity holdings from a variety of sources. These include annual reports, investment performance reports, and other information found on the websites of public pension plans or investment units associated with these funds. The data covers the largest pension plans or retirement systems in the majority of U.S. states. However, Arizona, Colorado, Nebraska, Nevada, Ohio, Tennessee, Utah, Virginia, and West Virginia are exceptions due to limited data availability. In total, the analysis includes 163 public pension plans.

To associate the VC funds listed in VentureXpert with those held by public pension plans, I use various adaptations of fund names. For example, a fund labeled 'Providence Equity Partners III' might correspond to versions like 'Providence Equity Prtnrs III', 'Providence Equity Ptrs III', 'Providence Eq Partners III', or 'Providence Eq Ptrs III'. Following this procedure, I manually search for the private equity funds'

[5]The Information on patent applications is gathered from both pre-grant publications and granted patents.

names that are unmatched to VC funds in VentureXpert. if available I note the identifier of each private equity fund reported by public pension plans. Upon concluding this step, I find that around 6.8% of VC funds count public pension funds among their LPs.

Other information on public pention plans is collected from the Public Plans Database (PPD) developed by the Center for Retirement Research at Boston College (CRR) and the MissionSquare Research Institute. The PPD is a comprehensive database that encompasses plan-level information from 2001 to 2022, covering 220 pension plans. Of these, 119 plans are administered at the state level, while 101 plans are administered locally. This sample represents approximately 95 percent of public pension membership and assets across the nation, ensuring broad coverage and representation. The majority of the public pension funds on my list as LPs coincide with the plans included in PPD's coverage.

1.1.4 Other Data Sources

In order to gauge the degree of connection between a startup's technology and public firms, I first identify patent applications that have been filed by these public companies. To pinpoint the public firm assignees of patents, I utilize the names of public firms obtained from both CRSP and Compustat.

I also incorporate data regarding inter-airport distances, sourced from the Bureau of Transportation Statistics. This information allows me to calculate the probable distances traveled between any two states by GPs and LPs. The dataset includes distances between all major airports, so to determine the likely travel distance that may be covered by GPs or LPs, I designate a specific airport for each state. To that end, I manually identify the largest airport in each state, except for California. For California, I opt to use San Francisco Airport, primarily due to its proximity to Silicon Valley. I list all the airports included in the research in the Online Appendix.

1.2 Start-Up Technologies and Exits

In this section, I demonstrate how the technologies and patents of startups can affect the likelihood of exits. Exit strategies can encompass initial public offerings (IPOs), acquisitions, mergers, or other means of monetizing or withdrawing from an investment. I specifically focus on IPOs and acquisitions.

In my sample, for startups whose last VC investment was before 2018, 11.5% of them exited through IPOs and 29.7% exited through acquisitions. While acquisitions occur more frequently than IPOs in the startup ecosystem, IPOs are generally recognized to yield the highest returns for venture investors. As a result, an IPO is often described as the most desired yet least common exit strategy for a portfolio venture.[6]

Given the prevalence of acquisitions as an exit strategy, I explore the potential influence of a startup's technology on its exit likelihood. I posit that startups whose innovations are closely related to technologies pioneered by public firms are more likely to be acquired. This is due to several strategic motivations of public companies. Public firms, in their quest to safeguard their market leadership, often find it necessary to acquire startups with similar technologies to mitigate potential market threats (Cunningham, Ederer, and Ma (2021), Kamepalli, Rajan, and Zingales (2020)).[7] Additionally, established firms may seek to integrate technologies that align closely with their own to achieve synergies, particularly in areas such as research and development, market expansion, and operational efficiency (Higgins and Rodriguez (2006), Makri, Hitt, and Lane (2010)).

To test the hypothesis, I create a metric to measure the similarity or compatibility of technologies to those of public firms. To that end, I utilize patent applications of startups. Specifically, I create a variable $PublicTechBase\_Avg$, which is the proportion of inventions by public firms in the patent references cited by startups before their exits. This is averaged across all of a venture's inventions during its tenure as a startup. This variable quantifies the degree to which a start-up relies on public firms' technologies for its technology and signifies the level of relationship between the start-up's technology and that of public firms.

Another metric that is closely related to the likelihood of exit is the number of patents granted to startups. First, holding patents increases the intellectual property value of the startup, which may relate to the fundamentals of the startup. For potential acquirers or investors, this can be a highly attractive asset, potentially leading to a higher valuation of the startup (Cotei and Farhat (2018), Gaule (2018), Farre-Mensa, Hegde, and Ljungqvist (2020)). Also, startups with strong patent portfolios

[6] Nahata (2008) use this fact to show top-tier VCs are most commonly linked to IPOs, mid-tier VCs to acquisitions, and the least respected VCs to unsuccessful exits. Moreover, according to Q2 2022 PitchBook-NVCA Venture Monitor report, an IPO is both the most preferred and the least common method of exiting a portfolio company. It shows that IPOs represent 71.4% of the total value of exits in 2021, but they are fewer in number. The most frequent type of exit is a takeover, representing approximately two-thirds of all deals. Link to the report.

[7] Similarly, Hoberg and Phillips (2010) explore the idea in the product market context. Acquisitions are more likely between companies that employ comparable product market terminology. Following transactions, stock returns increase when the involved companies use similar product market language. This trend is particularly strong in highly competitive product markets.

are often more successful in raising additional capital, which can be a stepping stone to a future exit (Hochberg, Serrano, and Ziedonis (2018)).

To assess if the number of patents granted to a startup serves as a predictor for exits, I introduce a variable, $Log(\#Patents+1)$, which is the logarithm of the number of patents granted to a startup prior to an exit.

I control for a variable, $Log(RefIntencity)$, measuring the degree to which a startup leverages the existing technology base. This variable is the logarithm (plus one) of the number of patent references cited by the startup before an exit, normalized by the number of patent applications they submitted prior to that exit. This control aims to mitigate the possibility of a mechanical relationship between the two variables above and startups operating in fields that inherently depend on past technology. In certain model specifications, I also introduce interaction terms between the startup's economic sector and its founding year as fixed effects, in order to account for variations in technological environments over time.

For startups with patents identified in Google Patent, I calculate the metrics $PublicTechBase\_Avg$ and $Log(RefIntensity)$. For those without patent data, I estimate these metrics using business descriptions from Refinitiv. I match startups based on business descriptions to derive mean values from those with patents, which are then used to estimate values for those without. The combined metrics, incorporating both actual and estimated values, are denoted as $PublicTechBase\_Avg\_Imp$ and $Log(RefIntensity)\_Imp$. Further details are in the Technical Appendix.

The dependent variables relate to the likelihood of a startup's exit. $I(Exit)$ is a binary indicator variable, assigned a value of one if a startup achieves a successful exit, and zero otherwise, with a successful exit being defined as either an acquisition or an IPO. Similarly, $I(Exit\_Acq)$ is a binary indicator where a value of one denotes a startup's exit via acquisition, and zero otherwise. Conversely, $I(Exit\_IPO)$ is set to one if the startup exits through an IPO, and zero otherwise.

With the variables, I examine the linear relationship between the technologies of startups and exit properties. **Table 2** reports the results. The sample comprises startups that received funding from VCs between 1980 and 2017, but their most recent investment occurred prior to 2018. The chosen sampling period ensures that startups have sufficient time to realize their outcomes following the final investment received from VCs.

I first examine the linear relationship with a sample of startups with patent applications in Panel A. In the first three regressions without fixed effects, I find that a one-standard-deviation increase in $PublicTechBase$ (0.1838, unreported) corresponds to a 4.71 percentage point rise in the probability of exit. This increase represents approximately 11.45% more chances of successful exits compared to the unconditional probability of successful exits. The increase in the probability of exit is driven by the increased likelihood of exits through acquisitions. An increase of one standard deviation in $PublicTechBase$ is associated with a 4.98 percentage point increase in the probability of exit through acquisitions. This increment represents roughly a 16.77% higher likelihood of successful exits compared to the mean value of the acquisition rate. On the other hand, $PublicTechBase$ does not show any relationship with the IPO. The findings, controlling for fixed effects of economic sector and venture founding year combinations, yield similar inferences.

In Panel B, I examine all startups, both those with patent applications and those without, using imputed values of $PublicTechBase\_Avg$ and $Log(RefIntensity)$. The results are similar to those in Panel A. In the first three regressions, which do not control for fixed effects, a one-standard-deviation increase in $PublicTechBase\_Avg\_Imp$ is linked to a 2.98 percentage point increase in exit likelihood. Contextually, this surge approximates to about 7.25% enhanced likelihood relative to the baseline probability of exits. Similarly, a one-standard-deviation increase in $PublicTechBase\_Avg\_Imp$ leads to a 4.09 percentage point rise in the likelihood of acquisitions, which corresponds to an approximate 13.77% greater chance of successful exits relative to the average acquisition rate. The coefficient of $PublicTechBase\_Avg\_Imp$ is negative and statistically significant in column (3), but the statistical significance disappears once I control for the fixed effects for the combinations of economic sectors of startups and their founding years.

The variable $Log(\#Patents + 1)$ shows a strong positive correlation with the likelihood of exits, both through acquisitions and IPOs. In the fixed-effects model outlined in Panel B, a one-standard-deviation increase in $Log(\#Patents+1)$ results in a 2.88 percentage point increase in the likelihood of being acquired. This translates to roughly a 9.70% higher chance of a successful exit compared to the average acquisition rate. Similarly, a one-standard-deviation increase in $Log(\#Patents + 1)$ leads to a 2.98 percentage point increase in the likelihood of IPOs, equating to approximately a 25.95% higher chance of a successful exit relative to the average IPO rate.

The analysis reveals that investing in startups with technologies similar to those of public companies increases the probability of exit through acquisition. While acquisitions are generally seen as less lucrative than IPOs, they offer a more certain pathway to a successful exit. Additionally, the presence of patents is a strong predictor of successful exits. However, firms with existing patents often already have higher valuations and multiple investors, presenting venture capitalists with potentially lower returns. As a result, venture capital managers are faced with a trade-off between risk and return. In the following section, I will explore how this dynamic affects venture capitalist decisions when public pension funds are involved as limited partners.

1.3 Effect of Public Pension Fund on VC Funds

In this section, I examine how the involvement of public pensions as investors in VC funds affects the investments these funds make. To analyze this effect, I employ a two-stage least squares model where the treatment is an investment from a public pension. In the subsequent subsection, I will outline the challenges that arise within this context and propose a research design to address them.

1.3.1 Public Pension Investors and Deal Characteristics: Linear Relationship

The General Partners (GPs) or a manager of a VC fund hold a significant degree of discretion and decision-making power when it comes to selecting which startups to fund and overseeing those startups. Typically, securing funds from Limited Partners (LPs) precedes the selection of startups by GPs. Therefore, by the time when GPs actively source potential investment opportunities, they already know who their investors are and what their investors want.

Let us use $Y_{i,n,s,t}$ to represent the variable that captures the decisions of a manager of a VC fund i on its nth investment, which was founded in a year t in a state s. These decisions of managers may encompass both the selection of startups for investment and the specific direction they steer these startups post-investment. The following model suggests that the presence of public pensions can influence such investment and guidance decisions.

$$Y_{i,n,s,t} = \beta_0 + \beta_1 I(PPFs)_{i,s,t} + X_{i,n,s,t}\gamma + \epsilon_{i,n,s,t}, \tag{1}$$

$I(PPFs)$ represents the presence of public pensions, which is set to one when a fund has a public pension fund as an investor and zero otherwise. X is a row vector of covariates, which captures the variations in the investment environments, and ϵ denotes the error term.

The dependent variable, $Y_{i,n,s,t}$, is chosen to capture the risk management practices of VCs, with an emphasis on two primary sets of aspects.

The first pertains to the kind of technologies pursued by startups. Technologies built upon those created by public firms suggest validation by these established entities. Moreover, as indicated by the analysis in the prior section, such technologies are strong predictors of acquisitions. As a result, I construct a measure that gauges the technological proximity of a startup, utilizing information available at the point when a VC fund manager is deciding on a deal.

For startups with patent data, $PublicTechBase\_Imp_{i,n}$ represents the proportion of patents from public firms that a venture cites in its most recent patent publication. For startups lacking patent information, the value is derived from similar startups, using their business descriptions and the year of investment as matching criteria. Consequently, all relevant information is available at the time VC funds make their investments. In a similar vein, I also tally the number of unique public firms referenced in a venture's citations.

The second suite of variables focuses on the deal stage. startups in later-stage deals offer more information on their prospective exit paths. Thus, startups at these stages encounter less uncertainty compared to those in the early stages. To capture this, I introduce a dummy variable for early-stage deals. Furthermore, VC funds may partake in deals with startups that have recently secured late-stage funding, as these are essentially vetted by other investors, indicating their worth for later-stage investments. I incorporate this by creating another dummy variable for investments in which a venture has raised late-round capital within the preceding two years.

In the regression models, I incorporate fixed effects for the combination of the fund focus - which generally refers to the predominant economic sectors of startups targeted by the funds - and the inception year to account for the evolving investment opportunities within each sector.[8] Additionally, I find that VC funds designated for early or late-stage investments do not exclusively invest in startups at those particular stages. Specifically, only 43% of the investments made by early-stage VC funds are in early-stage investments, loosely referring to seed or early-stage ventures, indicating that a substantial 57% are in later stages. Similarly, late-stage specified funds invest 19% in early stages, while generalist or balanced-stage funds invest 18.7% and 38.5%

[8]The vintage year of a VC fund, indicating its initial investment period, is critical for accurate comparisons. It accounts for different market and economic conditions impacting fund performance. VentureXpert lists a fund's founding year, which may not match its vintage year due to varied factors influencing the start of investments. To navigate this, my analysis controls for the founding year times fund's focus fixed effects, as funds established in the same year and sector often have similar investment timelines. In my main analyses on the effect of public pensions, I control for the combinations of founding year, fund focus, fund stage, age group, and size group to compare VC funds with similar characteristics as closely as possible.

in early stages, respectively. Consequently, to control for the propensity of VC funds to invest in certain stages of ventures, I also include fixed effects based on the specified stage focus of the funds.

Table 3 reports the results from OLS regressions. The results indicate statistically significant relationships between the presence of public pension investors and the investment characteristics associated with risk. Specifically, having public pensions as investors is significantly associated with increased portfolio startups' technological proximity to public firms and a rise in the number of unique public firms that startups cite in their patents. Moreover, there is a significant positive association between public pension presence and VCs' choices for later-stage deals. Also, there is a strong positive relationship with deals in which a venture has received financial support from other VCs in the venture's preceding late rounds.

1.3.2 Two-Stage Least Squares with Saturated Covariates

Public pension funds evaluate a range of factors before deciding on the VC funds in which they invest. Consequently, the presence of public pension funds in a VC fund can be seen as an aggregate outcome of the individual decisions made by these pension funds. The following equation illustrates the idea.

$$I(PPFs)_{i,s,t} = \phi_0 + Z_{i,s,t}\delta_1 + X_{i,s,t}\delta_1 + \nu_{i,s,t}, \tag{2}$$

Z is another row vector of variables that explain the number of Public pensions. ν is the error term orthogonal to Z.

The two error terms, ϵ, and ν, could potentially share common, unobserved factors. These factors might include the performance of other VC funds within the same investment company (Barber and Yasuda (2017)), the competency of managers (Kaplan and Schoar (2005), Korteweg and Sorensen (2017)), and the incentives presented to these managers (Gifford (1997), Abuzov, Gornall, and Strebulaev (2022)). Public pensions likely possess a thorough understanding of the performance of past funds and might consider it a significant factor when communicating with VC managers before investing in a fund. Furthermore, managers often transfer their distinct investment style to subsequent funds they oversee. If public pensions have specific requirements and identify a manager whose approach aligns with those needs, they are more inclined to invest in funds managed by that individual.

The issue of selection bias resulting from omitted variable bias is particularly pronounced because VC funds operate in an opaque industry. As such, there is a probable correlation between these error terms. Then, regressing Y on $I(PPFs)$ results in an inconsistent estimator as the asymptotic covariance between $I(PPFs)$ and ϵ in equation (1) does not converge to zero. In other words, the coefficient of $I(PPFs)$ is not interpreted as causal effects since the investments from public pension funds are not randomly assigned.

To address the selection bias, I formulate an instrumental variable that affects the choices of VC managers exclusively through the presence of public pension funds as LPs. The idea in its core is to capture the synchronicity between two crucial components: one is the timing when the VC managers are actively raising capital for a follow-on fund as found in the literature, and the other factor is the fund's location in areas where individual VC funds are likely to receive large investments from neighboring public pensions.

The first consideration centers on the timing when GPs seek fresh investments from LPs, a dynamic deeply rooted in the institutional fabric of the VC industry. According to Barber and Yasuda (2017), it is customary for VC fund managers to initiate a new fund between the third and sixth year of a current fund's lifecycle. This timing aligns with when managers are poised to highlight the key investments made during the early phases of the ongoing fund. With this backdrop, I propose that as VC firms move beyond the primary investment stage and commence showcasing their fund's achievements to potential investors, they have a heightened likelihood of securing fresh investments for their new funds.

The second component, the location where individual VC funds can expect public pension money, encompasses two intertwined facets: the geographical closeness of public pensions to VC funds and the pensions' plan toward investing in private equity.

Public pensions may consider geographical proximity when investing in VC funds. First, the costs associated with supervising portfolio managers diminish when those managers are closer to the pension investors. The reasoning stems from informational benefits and trimmed travel expenses (Coval and Moskowitz (2001), Tian (2011), Ellis, Madureira, and Underwood (2020)). Since pension funds disclose costs associated with their asset management, they are motivated to pare down expenses linked to liaising with these managers. Moreover, investing in VC funds in nearby areas enhances support for in-state startups, since proximity also induces VCs to invest in startups closer to them to more easily acquire soft information about these startups and conduct post-investment monitoring. Public pensions, in particular, are driven by an incentive to invigorate their regional economies, aiming to attract new taxpayers to their states.

In Appendix B, I analyze interstate investment patterns, tracing investments from pension funds in a given state to VC funds in its N_{th} nearest neighboring state. The findings suggest that the model's goodness of fit is highest when focusing on states within the top five closest states, including the home state.

An explanation of this finding is the uneven distribution of VC firms across states. For example, the Providence Employee Retirement System of Rhode Island predominantly invests in nearby VC hubs like Boston, Chicago, and New York, likely due to Rhode Island's limited VC landscape. This finding does not clash with the finding of Hochberg and Rauh (2013), which highlights a public pension fund bias towards in-state VC investments. My analysis indicates that the home-state dummy variable registers the largest coefficient at 0.075, implying that VC funds within the same state are most likely to receive investments from local pensions. However, given the geographical clustering of VC funds in certain areas, those situated near the state of the pension funds - not just within the same state - might also experience an increased likelihood of securing capital from them.

Another facet of the second component concerns the discrepancy between a public plan's target private equity allocation and its actual private equity allocation. I posit that when this gap is large, the plan is more likely to channel additional assets into private equity, including venture capital.

Combining the two aspects of the second component allows us to identify areas where VC funds might anticipate substantial inflows from neighboring public pensions. These areas are states situated near those where public pensions are preparing to invest in private equity.

Given the two components—the first being the timing when VC managers are actively seeking capital for a subsequent fund, and the second being the proximity to public pensions aiming to increase their private equity allocations—it is plausible that these factors are quasi-random in relation to each other. Consequently, the instrumental variable is the intersection of these two components. It captures the establishment of a VC fund coinciding with the typical VC fundraising cycle and its location in a state with a favorable probability of receiving public pension funding during the capital-raising phase. This variable indicates an increased likelihood of a VC fund securing capital from public pensions simply because it is in the stage of raising capital for its next fund while nearby pensions are concurrently looking to allocate resources to private equity.

To initiate the variable construction, I compute the gap between the target and actual private equity allocation of a public pension plan, expressed in dollar terms.

By summing up these figures across all retirement plans within the state, I arrive at the total dollar amount that could be allocated to private equity for each year by the state. The amount calculated for a state s in a given year t is denoted as $ExpectedPEinflow_{s,t}$.

Subsequently, I identify the five geographically nearest states to a specific state. This approach is grounded in the empirical findings in Appendix B that VC funds situated in these proximate states have a higher chance of securing funds collectively compared to those in more distant states.

To determine the five closest states to any given state, I use inter-airport distances as a proxy for potential travel distances between states for VC managers and pension fund directors. This method involves assigning a major airport to each state as a reference point, with San Francisco Airport selected for California given its proximity to Silicon Valley. A detailed list of the chosen airports for each state is available in the Technical Appendix.

Following that, I assume that the additional funding available to VC firms in neighboring states, arising from the difference between the targeted and actual private equity allocations in a state, is disproportional to the number of VC funds established in those close states in the last three years. Let us represent the number of VC funds founded in a state k in the three years preceding year t as $Num\_Funds_{k,t}$. Then, the potential surge in new investments available to each individual VC fund in nearby states, due to the PE allocation discrepancy in state s for a specific year, can be quantified by the variable $flowPerFund_{s,t}$.

$$flowPerFund_{s,t} = \frac{ExpectedPEinflow_{s,t}}{\sum_{n \in N_s} Num\_Funds_{n,t}} \tag{3}$$

where N_s is a set comprising the five states closest to a given state s, including the state s. The variable $flowPerFund_{s,t}$ is likely to be proportional to the actual capital that could be allocated to each VC fund in close proximity to the location of the pension funds.

To identify the state-year observations with sufficient capital available for each VC fund, I rank the variable $flowPerFund_{s,t}$ across all state-year observations for the sample period spanning 2001 to 2020, sorting them in descending order. From this sorted list, I select the upper quantile to represent cases where public pension funds in a given state s during a specific year t have substantial assets that could be allocated to nearby individual VC funds. I designate the collection of state-year pairs that correspond to these instances as set $\boldsymbol{P} = \{(s,t) \in \text{The upper quantile of } flowPerFund_{s,t}\}$.

Based on the set, I create an indicator variable ($PensionFlowTiming_{s,t}$) for each state s and year t where VC funds might have a good chance of getting funds from public pension funds because the funds are located in the state s and in a year t. Specifically,

$$PensionFlowTiming_{s,t} = \begin{cases} 1 & \text{if } (s,t) \in \boldsymbol{P}, \text{ or } (s,t) \in \{(s,t)\ ;\ (k,t) \in \boldsymbol{P},\ s \in N_k, \} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

where N_k is a set comprising the five states closest to a given state k.

Next, I incorporate the timing of VC managers seeking new investments from LPs. Specifically, I assume that VCs actively seek new funding from investors between three to six years after the establishment of a fund (Barber and Yasuda (2017)).[9] For each VC fund with a given founding year t, I create an indicator variable ($VCFundTiming_{i,s,t}$), which will serve to indicate that the fund's creation was motivated by the timing to raise capital, which was solely driven by the lifecycle of their preceding funds. Specifically, $VCFundTiming_{i,s,t}$ is assigned a value of one if there was a VC fund established within the same VC firm between three to six years prior to the inception of the focal fund i; otherwise, it takes a value of zero.

Finally, I construct an instrumental variable that captures the synchronicity between two components. When a VC fund i within a certain state s seeks to secure funds one year prior to its inception year t, there is a heightened likelihood that it will receive investments from public pension funds due to the fund's close proximity to those public pensions with heightened interest in allocating capital to the private equity sector. The instrumental variable $Pension\_VC\_TimingSync_{i,s,t}$, which is created for each VC fund i with a founding year t in a state s is described by

$$Pension\_VC\_TimingSync_{i,s,t} = PensionFlowTiming_{s,t-1} \times VCFundTiming_{i,s,t} \tag{5}$$

[9] Using Preqin data, Barber and Yasuda (2017) utilize the vintage year of a VC fund as a reference point to identify the period when managers are likely to engage in capital raising for a follow-on fund. I use the founding year of a VC fund, as provided by VentureXpert, to serve as a proxy for the vintage year, since the gap between the founding and vintage years is typically just a few months to a couple of years. To address the possibility that the assumption may not hold, I run reduced form regressions using another instrumental variable created from the first year of investments showing up in VentureXpert as the vintage year. I find that the reduced form regressions shows similar results to the reduced form regression using the instrumental variable in the paper. The results are available in the Internet Appendix.

The variable indicates that a VC fund i had an increased probability of getting investments from public pensions when established in year t. This increased likelihood primarily arises from the synchrony between the fund's fundraising efforts in the preceding year and the investment interests of nearby public pension funds in private equity during that same year.

I use two-stage least squares (TSLS) to estimate the parameters. Angrist and Imbens (1995) show that TSLS can be applied to identify a weighted average of treatment effects along with properly defined weighting functions. The key approach described in their paper is to construct a model specification in a way that causal response weighting functions are empirically determined (see Theorem 3 in Angrist and Imbens (1995)). The recent work of Blandhol et al. (2022) highlights that TSLS may fail to achieve a local average treatment effect interpretation without such model specification. Therefore, I use a TSLS specification that controls for covariates by including dummies for all possible combinations of the values of covariates as described by Angrist and Imbens (1995).

Specifically, in the first-stage regression, $I(PPFs)$ is regressed on dummies representing each unique combination of covariates and the instrumental variable. The predicted value from the regression, $I(\widehat{PPFs})$, is then incorporated into the second-stage regression as a substitute for $I(PPFs)$. This second-stage regression also includes indicator variables corresponding to each unique combination of covariate values.

1.3.3 Assessment of the Instrumental Variable

1.3.3.1 *Relevance Condition*

Before presenting the findings from the second stage regressions, I conduct an assessment of the instrument variable's relevance to the presence of public pension funds as LPs in a VC fund. In this analysis, I focus on a sample comprising U.S. VC funds with a vintage year ranging from 2001 to 2021. **Table 4** shows the relationship between the instrumental variable, $Pension\_VC\_TimingSync$, and the main variable, $I(PPFs)$. The coefficients in regressions with and without fixed effects are statistically significant at the 1% significance level. In Column (3), a VC fund with a value of one for the instrumental variable is 13 percentage points more likely to secure capital from pension funds. This represents a 175% increase compared to the average likelihood of attracting public pensions as investors.

In the setting described by Angrist and Imbens (1995), incorporating covariates into the causal model necessitates the inclusion of a comprehensive set of dummy variables to fully saturate the combinations of covariate values. In this case, a sufficient number of observations is essential for reliable estimation. Moreover, as the main hypothesis examines the decisions of VC managers at the investment level, the analysis focuses on investment-level data. In this approach, the unit of observation is defined as a pair consisting of a startup in a specific funding round and a VC fund investing in that round.

I examine the relationship between the instrumental variable and $I(PPFs)$ using a sample of investments made by VC funds over the same period. I find that the instrumental variable retains its explanatory power even after accounting for the age and size of the VC firm - where the size is measured by the average deal number of the VC firm's past funds showing up in VentureXpert. In Panel B, Column (3), for a fund within a specific age-size category of its VC firm, focusing primarily on startups in a particular stage and economic sector in a given year, the instrumental variable is associated with an average increase of 14.6 percentage points in the probability of receiving funding from public pensions.

The findings underscore the significant explanatory power of the temporal alignment between nearby public pension funds increasing their allocations to private equity and VC fundraising activities in accounting for the presence of these pension funds as LPs in VC funds.

1.3.3.2 *Exclusion Condition*

The IV satisfies the exclusion restriction, particularly concerning the selection of public pensions. Firstly, the IV is unlikely to be correlated with the risk attributes of preceding funds or with managers' risk preferences, which are elements public pensions might consider. This is because a VC's fundraising cycle is generally independent of these factors. Suppose that there are two VC funds in a state that has a high likelihood of receiving pension money. Let us say the risk preferences of the managers of the two funds are the same and what the public pension is looking for. The IV only lights up for the fund that has an earlier fund that was created three to six years ago. Hence, the IV does not correlate with what public pensions may seek in terms of the risk profiles of past funds or the risk preferences of managers.

Secondly, public pensions cannot easily anticipate that certain types of VC funds will relocate to benefit from increased capital flow. The challenges associated with launching a VC fund in a new state are considerable, due to competitive dynamics established by local incumbents (Hochberg, Ljungqvist, and Lu (2010)), the localized nature of information transmission (Sorenson and Stuart (2001)), and regional biases in VC investments (Cumming and Dai (2010)). Moreover, in Appendix C where I provide summary statistics of the IV by state, there is considerable variation within each state regarding the IV. The substantial variation in the IV within individual states implies that it is challenging for VC managers to use this information to actively select locations where they can benefit from investments from public pensions. In other words, public pensions cannot expect that a specific type of VC fund will choose states near their own when they decide to shift their focus towards private equity.

Third, the IV does not represent specific states. Appendix C shows that there are no discernible patterns indicating which states are more likely to be regions where individual VC funds raised capital when there was a high likelihood of a large inflow from public pensions. States experiencing high levels of VC activity, such as California, New York, and Massachusetts, do not exhibit a high average value for the IV. Also, political leanings do not appear to account for the mean value of the IV in a state. Therefore, the IV does not represent a style of VC funds flourishing in a certain state.

1.3.4 Effects of Public Pension Funds on Investment Characteristics

Table 5 presents the effect of the presence of public pension funds on VC investment attributes associated with risk, obtained from the second-stage regressions. In the first-stage regressions, $I(PPFs)$ is regressed on interaction terms between the instrumental variable and a comprehensive set of indicator variables, each representing a unique combination of the values for the covariates specified in the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$, obtained from the first-stage regression, along with a set of indicator variables that capture every unique combination of covariates from the specified columns.

I introduce control variables in a phased manner. Initially, I consider only the fund focus, vintage year, and fund stage. Subsequently, I incorporate the age group and size group. The age group categorizes VC firms into three equal segments based on the firm's age, with divisions made annually. Similarly, the size group divides VC firms into three equal segments based on the average number of deals made by their previous funds, also categorized annually. I opt for defining these age and size groups rather than generating dummy variables for each unique value of VC firm age and size. This approach is designed to ensure a sufficient number of observations for each unique combination of control variable values.

The dependent variables remain the same as those in Table 3. **Table 5** presents the results of the second-stage regressions. These findings align with the hypothesis that VC managers construct less risky portfolios when their investors include public pension funds.

Firstly, the involvement of public pension funds is linked to a heightened level of technological alignment between portfolio startups and public firms. This is likely to increase the probability of successful exits via acquisitions for these startups. The results in Column (3) show that the average effect of public pension presence—affected by the VC fundraising cycle and geographical location, after controlling for given covariates—is 0.025. This represents an increase of over ten percent from the variable's mean value.

Secondly, VCs with pension fund investors tend to favor startups that already possess patents. The more a startup has patents the more likely it is invested my VC funds backed by public pensions. This observation correlates with an increased likelihood of a successful exit, either through acquisition or an IPO.

Thirdly, such VCs are more inclined to steer clear of early-stage deals, opting instead for later-stage investments. This behavior aligns with the notion that late-stage investments are not only more appealing because the associated startups more frequently go public, but also because they do so more swiftly, as indicated by Sorensen (2004).

Finally, VCs are more inclined to participate in deals that have already received validation from other investors during later investment rounds. The absolute magnitude of the coefficient is larger than the coefficient of the same variable when the dependent variable is the early-stage dummy. For example, the absolute magnitude of the coefficient in column (10) is larger than that in column (7). This suggests that among late-stage deals, those that have garnered validation from other funds in the late rounds are more likely to attract investment from VC funds with public pension investors.

1.3.5 Effects of Public Pension Funds on Startup Exits

After noting that VC funds with pension fund investors tend to select deals with a higher likelihood of successful exits, albeit with potentially lower rates of return, I proceed to explore whether this less risky investment strategy also carries over to the funds' exit strategies.

Table 6 reports the results from the second stage regressions with exit rates as the dependent variable. The analysis suggests that VC funds with public pensions have a higher likelihood of successful exits. The average impact of public pensions' presence, influenced by the VC fundraising cycle and location, on exit likelihood, as shown in Column (9) is a staggering 0.144. This number is approximately half of the unconditional probability of exits.

This finding corroborates the hypothesis that VC managers tactically steer their portfolio companies to enhance payoff probabilities, aligning with the outlays typical public pensions expect. This behavior could also stem from venture capital firms' reactions to the legal context governing public investors, which affects the confidentiality of private equity partnerships (Abuzov, Gornall, and Strebulaev (2022)). Specifically, VCs may avoid excessive risk-taking in funds with public pension investors due to performance disclosures influenced by such legal stipulations.

In summary, the increase in overall exit rates implies that VC managers construct and manage their portfolios in a way that complies with the prudent asset management criteria commonly associated with public pension funds.

1.3.6 Effects of Public Pension Funds on the Time to Exit in VC Investments

The influence of public pension funds includes VC managers favoring startups with many patents and late-stage deals, especially those that have recently secured late-round funding. This strategic choice likely impacts the time it takes for portfolio startups to exit. Specifically, startups in later funding stages generally exit more quickly than those in earlier rounds.

Additionally, the selection of startups with a higher likelihood of successful exits - based on their technological link to existing public firms - tends to expedite the exit process. It is because VC managers may pressure their startups to exit through acquisitions when there are potential acquirers (Masulis and Nahata (2011)). When a startup's technology is linked to existing public firms, there is an increased likelihood of finding potential acquirers. If VC managers recognize that their investors prefer regular payoffs, they may steer their startups toward exits via acquisition whenever a suitable acquirer emerges. This is in contrast to other funds that may not prioritize a timely exit, giving enough time for their startups to mature enough to maximize their value.

To delve deeper into this aspect, I introduce a variable to quantify the duration required for an investment to reach a successful exit after a VC fund was created. This duration is measured in years and spans from the founding year of the VC fund to the date of a successful investment exit where exits are confined to those that occur within a 15-year window following the fund's inception.

Table 7 reports the results from the second stage regressions with a sample of investments made by VC funds that ultimately achieved successful exits. The results suggest that VC managers, when backed by public pension funds, strategically construct and manage their portfolios to facilitate more expedient exits within the fund's lifecycle. In Column (3), which considers all exit types, the regression shows that the average effect of public pension presence, influenced by the VC fund-raising cycle and location, reduces the time to exit by 1.658 years. The t-statistics and R^2 values are higher in the sample of acquisitions, suggesting that the process of exiting through acquisitions is more straightforward compared to IPOs, which are more dependent on favorable macroeconomic conditions. The findings indicate that VC funds receiving investments from public pensions tend to realize exits more swiftly.

1.4 Other Results

In the following sections, I present analyses that support the main findings or enhance the comprehension of them. First, I examine the relationship between the presence of public pensions in a VC firm and the risk-taking across all investments at the VC firm level. Second, I examine whether investment-level risk metrics are associated with the returns in a way that is consistent with the hypothesis, i.e., less risky investments lead to lower returns. Third, I investigate the investment choices of public pension investors, differentiating between first-time investors in a VC firm and repeat investors with prior experience in the firm's funds. I also consider the financial health of public pensions as a factor influencing their decisions to invest in VC funds with specific risk profiles. Lastly, I explore whether the financial condition of public pensions plays a role in the fund management strategies employed by VC managers.

1.4.1 Presence of Public Pensions and VC Firms' Risk-Taking

The participation of public pensions tends to reduce the overall risk of a VC fund's portfolio. However, the VC industry is typically marked by its chase for high-risk, high-reward opportunities. If public pensions push VC funds towards a less risky investment style, clients who seek those occasional, yet highly lucrative, investments might be sidelined. In this section, I explore the possibility of public pension funds crowding out these investors with different financial goals.

I construct an indicator variable, $I(PPFs\ in\ VC\ Firm)$, to represent the presence of public pensions in a VC firm. This variable is assigned a value of one if a VC firm has initiated a VC fund with public pensions within the past decade. Using a ten-year timeframe is a conservative approach to infer the involvement of public pensions. While a VC fund can exceed a lifespan of 10 years due to time to liquidate their investments for reasons like the inherent nature of their investments, prevailing market conditions, and strategic choices to optimize returns for the investors, they are predominantly active during the initial 10 years of their funds.

I examine an investment risk metric against the indicator variable denoting the presence of public pensions as investors in a VC firm. VC firm fixed effects are included to explore within-firm variation. For each risk metric, there are two model specifications, based on whether a dummy variable indicating a particular VC fund has public pension investors ($I(PPFs)$) is included. In this specification, the coefficient of $I(PPFs)$ signifies the difference between VC funds supported by public pensions and those that are not within the same VC firm.

Table 8 presents the results. First, in the regressions excluding the dummy for VC funds with public pension investors, for each investment-level characteristic, the coefficient of $I(PPFs; in; VC; Firm)$ is not statistically significant except for the number of patents. Moreover, the direction of the coefficients is contrary to what we observe for $I(PPFs)$. The findings imply that the presence of public pensions in a VC firm does not play a significant role in changing the overall risk management style of the firm.

In model specifications where $I(PPFs)$ is included, the coefficients for $I(PPFs)$ align with earlier findings. Funds supported by VC funds backed by public pensions tend to favor less risky investments. The inclusion of firm-fixed effects suggests that these funds are less risky compared to other funds within the same firm. Furthermore, the coefficients for $I(PPFs; in; VC; Firm)$ are the opposite of those for $I(PPFs)$, indicating that the overall investment characteristics of VC funds not invested by public pensions lean towards riskier deals in times when public pensions are VC firms' investors.

The findings from this section suggest that the presence of public pensions does not necessarily drive out other investors. Instead, the results indicate that VC firms redistribute risk, channeling less of it towards funds supported by public pensions and more towards those that are not. Consequently, the overall risk management approach of a VC firm appears to remain constant, or potentially riskier, regardless of the involvement of public pensions in their portfolio.

1.4.2 Investment Characteristics and Round-to-Round Returns

The four investment risk metrics are associated with a likelihood of exits. However, investments that indicate a higher probability of exit do not necessarily yield higher returns. In this section, I explore whether these four risk metrics correlate with the magnitude of returns in a way that is consistent with the risk-return trade-off.

For return computations, valuation data is essential. Such valuations become evident when a firm either goes public, obtains new financing, or gets acquired. In this framework, I assess investment-level returns by calculating round-to-round returns. These returns represent the potential gains from investing in one funding round and holding until the subsequent successful funding round or exit.

I start with all investment observations backed by U.S. VC funds from 1980 to 2022. Out of 574,382 investments that are backed by VC funds, I have data for 83,015 investments where the valuation is observed. However, not all observations are strictly VC deals; some of them are categorized as debt or LBOs. Therefore, I concentrate on round-to-round returns where two consecutive rounds are purely VC investments.

Additionally, I enhance the dataset by including startups with available exit values. For acquisitions, I utilize the exit values provided by the SDC database. For IPOs, I require observations to have data on the offering price, the number of shares issued, and the proceeds. I estimate the number of shares issued using the number of common shares outstanding, as found on CRSP, at the earliest available instance. I determine the valuation of IPOs based on the method described by Cochrane (2005).

For total round equity, I have data on 474,728 observations. Due to the scarcity of direct valuation observations and the relatively abundant data on total round equity and exit values, I employ a method to estimate valuations for intermediate funding rounds between two data points when the intermediate investments lack direct valuation data.

Specifically, I employ linear interpolation based on the time between rounds of funding where valuation is observed. For instance, if there is a funding round missing valuation data between two rounds with known valuations - say, the startup was valued at V_1, K months prior, and V_2, Q months later - the interpolated post-money valuation for that missing round would be calculated as $K/(K+Q) \times (V_2 - V_1) + V1$. I use post-money valuation for V_1 and pre-money valuation for V_2.

I compute round-to-round returns by taking into account the interests of existing investors and capturing the portion of the valuation that goes to the VC fund investing in the round. Specifically, the round-to-round return is calculated as the ratio of the post-money valuation in one funding round to the pre-money valuation in the subsequent funding round minus one. I treat all investments as common stock. I treat each investment as distinct if a VC fund invests in the same startup on multiple occasions.

From the steps above, I find 35,092 investments with available returns. Yet, when

I limit the observations to those from funds established after 2000, for which data on public pension participation is available, the number of observations drops to 6,700. Given this, rather than running regressions with returns and the indicator for public pensions' presence, I opt to investigate the relationship between investment risk metrics and round-to-round returns. In the online appendix, I present findings indicating that the presence of public pension investors corresponds to a decrease in returns by 0.481 without accounting for any fixed effects.

Table 9 presents the results from OLS regressions where the dependent variable is round-to-round returns, and the independent variables are investment-level risk metrics. Consistent with our expectations, the level of connection to public firms' technology, the number of patents, and the indicator variable for recent late-round funding are all linked with lower returns. For the indicator variable for early-stage investments, the round-to-round returns are higher. Except for the linkage to public firms' technology, all three variables exhibit a statistically significant relationship with returns at the 1 percent significance level.

The magnitudes of the coefficients align with existing literature. For the indicator denoting early-stage investments, the return increases by 0.505, whereas the indicator for late rounds with recent late-round financing corresponds to a reduction in return by 0.443. Considering that the standard deviation of returns from 1987 to 2000 was 107% (Cochrane (2005)), these estimates appear reasonable.

Overall, the analysis suggests that VC managers' investment-level risk decisions lean towards strategies that offer a higher likelihood of payoff, albeit with low returns when public pensions invest in their funds. This highlights the notion that the mediocre returns seen in public pensions' VC investments may not solely be a disadvantage. Instead, these moderated returns might be a conscious decision by VC managers to ensure greater certainty of payoffs. This raises the question: Is this cautious approach in line with what public pensions anticipate from their asset managers? The following section explores whether this conservative risk stance enhances the likelihood of drawing public pensions to the VC firm.

1.4.3 Previous Funds' Characteristics and Public Pension Funds' Decisions

In this section, I explore the relationship between the participation of public pension funds in venture capital (VC) funds and the attributes of prior funds within the same VC firm. The track record of previous funds serves as valuable information for investors. There is empirical evidence supporting the idea. For instance, VC fundraising largely depends on the track record of the individual VC firm (Gompers and Lerner (1999), Kaplan and Schoar (2005)). Barber and Yasuda (2017) show that interim performance significantly influences fundraising outcomes. Hence, if investors base their decisions on the track records of VC firms, public pension investors might also assess the risk profile of these firms by examining their historical investment patterns.

The studies on why public pensions invest in alternative assets suggest a quest for higher yields due to underfunded objectives (Pennacchi and Rastad (2011), Mohan and Zhang (2014), Lu et al. (2019), Andonov, Bauer, and Cremers (2017)). In this context, the prudent risk management of VC funds might not align with what public pensions desire. In this section, I assess whether such public pensions favor VC funds that tend to structure and manage portfolios in a manner that amplifies investment risk, and in turn rewards, within the venture capital domain.

I focus on VC funds that had investment records from earlier VC funds, specifically those investments made three to six years prior to their own inception, inclusively. Next, I compute various metrics that serve as proxies for the characteristics of these earlier funds. The variable $PublicTechBase\_Past$ represents the average of $PublicTechBase\_Imp$ values from startups that previous funds invested in, within a three to six-year span before the focal fund was established. When none of the previous investments feature startups with the variable $PublicTechBase\_Imp$, I presume their technological connection to the public firm to be zero, aiming to boost the count of the funds. In analyses without this substitution, the coefficients of $PublicTechBase\_Past$ prove to be significant without fixed effects, yet they become insignificant when fixed effects are introduced. In the same vein, $log(\#Patents)\_Past$ is the mean value of the natural logarithm of the total patent count (incremented by one) from these earlier investments. The term $R(Early\_Inv)\_Past$ represents the fraction of early-stage investments made by these preceding funds in that same three to six-year period. Similarly, $R(Recent\_Late\_Round)\_Past$ denotes the percentage of later-stage investments that had another late round in the preceding two years, carried out by these older funds. Lastly, $Avg(Log(\#Deals))$ calculates the average number of investments made by these earlier funds.

Table 10 reports the results from OLS regressions where the dependent variable is an indicator variable that takes a value of one if the fund has a public pension as an investor and zero otherwise. The regression results reveal which characteristics of past funds correlate with investments from public pensions in the current fund and indicate the direction of these correlations.

Firstly, the average number of investments undertaken by preceding funds within the same VC firm positively influences the participation of public pensions in the VC fund. The conventional diversification metric, which emphasizes incorporating numerous investment assets into a portfolio, seems to be a criterion that public pensions consider when evaluating their investment options in VC funds. Moreover, in the prior analysis presented in **Table 9**, the variable positively correlates with returns. This suggests that a VC firm's average number of investments, closely associated with the size of the VC firm, also indicates the VC firm's capacity to find and invest in lucrative deals, and public pensions appear to value this attribute in their decision-making.

Turning to our four risk metrics, I find that the coefficient of $PublicTechBase\_Past$ is positive and statistically significant. Additionally, public pensions tend to select VC funds in firms where past VC funds have invested more heavily in startups with numerous patents. Lastly, public pension funds are more inclined to invest in VC firms whose past funds have prioritized later-stage deals, especially those with a history of late-round funding.

Overall, VC funds affiliated with firms that have a history of prioritizing lower-risk investments are more likely to attract public pension clients. However, it still remains unclear whether numerous public pensions opt for less risky funds or if it is only a specific pension that chooses such low-risk options. Also, the results do not tell whether these outcomes stem from public pensions actively seeking less risky funds or whether they discover a risk preference based on prior experiences with the funds managed conservatively. In the following section, I sharpen the results by distinguishing the selections made by public pensions both prior to and after investing in a VC firm.

1.4.4 Previous Funds' Characteristics and Public Pensions' Decisions

In this section, I aim to examine the preferences of public pensions amidst the myriad of VC fund options available to them. To achieve this, I assemble a dataset comprising unique pairings of public pension investment entities and VC funds available for investment at the time of the pension's decision-making. A VC fund is considered accessible to public pension investors if it was established in the same year, or a year before or after when the public pension made an investment in a VC fund. Also, the VC fund was located within the five states closest to the public pension. Additionally, these funds must have historical investment records from previous VC funds spanning from three to six years before their inception.

I use a linear probability model that includes fixed effects for pension investment units times investment years, fully capturing the pairing of public pension investment units with available VC funds for each year. These fixed effects are further intersected by the specific focus or sector of the VC funds. In essence, I analyze the choices made by a public pension in a particular year, given a specific set of investment options within a defined sector.

The investment attributes of previous funds are averaged across these funds, resulting in variables appended with the suffix *_Past*. Additionally, I create an indicator variable, *Past Client*, which is set to one if the public pension has previously been a client of the VC firm, and zero otherwise. This variable is designed to distinguish any influence stemming from the public pension's previous experience with the VC firm when evaluating the style of past funds.

Table 11 presents the findings. Initially, the regression outputs without the *Past Client* term yield conflicting outcomes. Regarding the technological orientation of startups, individual public pensions appear to lean towards riskier ventures. Yet, for the remaining risk metrics, the results align with the notion that public pensions also favor less risky deals.

Nevertheless, upon breaking down the effects according to whether it is a first-time investment in a VC firm or a subsequent investment post an initial association with a fund from that VC firm, a pattern emerges. The inclination toward less risky investments is primarily attributable to those who have previously observed less risky investments with the VC firm. As demonstrated in Column (2), public pensions that are investing in a VC firm for the first time seem to favor more risky startups. Conversely, those pensions that have invested in a VC fund with a track record of investments in less risky startups tend to gravitate towards another VC fund under the same VC firm's umbrella. The magnitude of the coefficient for the interaction term is larger, suggesting that the influence of past interactions with the VC firm is considerable in shaping investment choices.

For the remaining investment-level risk measures, the inference is the same. The tendency of public pensions to favor less risky VC investments is predominantly influenced by their past interactions with VC funds that exhibited their cautious investment strategies. The magnitude of the coefficients for the interaction terms is over eight times larger across all risk metrics.

This observation leads to several key takeaways. Firstly, public pensions appear content with the outcomes produced by a conservatively managed portfolio. Such a risk approach by a VC fund likely delivers returns that are both timely and adequate. Secondly, it seems that initially, public pensions may not have a pronounced risk preference. It is the subsequent experiences with VC funds, which adopt a cautious risk management strategy, that steer public pensions to maintain their association with that VC firm. Lastly, if the overarching goal of curating a less risky portfolio is to foster and preserve relationships with public pension investors, the evidence suggests that VC managers are indeed adept at achieving this aim.

1.4.5 Underfunded Public Pensions' Decisions on VC Funds

Moving forward, I further investigate whether public pensions' financial status affects their choices of VC funds. A recent study by Mittal (2022) demonstrates that underfunded public pensions tend to choose private equity GPs of lower quality, as indicated by size, resulting in low returns. In a similar vein, I examine whether public pension funds' decisions vary based on their financial health. In this section, I focus only on VC funds that are invested by public pension investors. This narrows the scope to examine the investment decisions of financially troubled public pensions that might differ from those that are financially stable.

The funded status and returns data come from the Public Plans Data database. Although the database does not cover every public pension in my sample, it claims that its data, covering fiscal years 2001-2022, represents 95% of pension assets and members at the state/local level in the U.S.

I employ two metrics as indicators of a public pension's financial health. Firstly, I utilize the funded ratio determined using the traditional GASB 25 standards. In accordance with GASB 25, this ratio is computed by dividing the actuarial assets by the actuarial liability. Secondly, I calculate the difference between the assumed rate of returns and the actual investment returns over the past three years. Under GASB reporting for public pension plans, the long-term investment return assumption is used to discount actuarial liabilities. The gap between the assumed rate and the realized investment return over the past three years serves as a proxy for the pace at which the pension plan needs to keep up with the assumed rate of return.

I define underfunded pensions as those whose reported funded ratio is below 0.9.[10] Also, I consider the gap between the assumed rate of returns and past realized returns to be large if it exceeds 0.03.[11]

[10] I also examined the funded ratio below one. I found that previous fund characteristics do not have a statistically significant impact on the participation of underfunded pensions defined by this particular criterion.

[11] I also explored instances where the gap is positive (greater than zero). My findings indicate that past fund characteristics do not significantly influence the participation of pensions with a positive

Table 12 reports the results. In the regression analyses, the dependent variable is an indicator representing the participation of financially unhealthy public pensions in the VC fund based on specific criteria. The variable takes a value of one if there is at least one public pension that is defined to be financially unstable, i.e. those that are underfunded or not keeping up with the assumed rate of returns, zero otherwise. The independent variables that represent the investment characteristics of past funds are the same as those in **Table 10**.

The findings indicate that there seems to be no significant difference in the way public pensions of varying financial health assess the past investments of VC funds. However, regarding the signs of the coefficients for the four risk metrics, pensions appear to favor VC funds whose preceding funds have chosen less risky investments. This preference might explain why underfunded public pensions often experience lower returns (Mittal (2022)). It is conceivable that the returns they receive, although seemingly modest, may not be as low when considered in terms of risk-adjusted returns. Essentially, they might be selecting investments with lower risk profiles, which naturally lead to lower expected returns.

This section sheds light on the determinants that financially challenged public pensions use to choose VC funds, drawing on the historical performance of those funds. Nonetheless, the findings in this section primarily pertain to a comparison among VC funds selected by public pensions. The question remains open as to whether these determinants also guide the decision-making of individual public pension investors when presented with various fund options. To address this gap, the following section delves into a broader analysis of individual pension plans' investment decisions, encompassing both selected and non-selected VC funds.

1.4.6 Previous Funds' Characteristics and Public Pensions' Financial Health

I examine whether and how the financial health of public pensions explains their selection of VC funds from the investment options available at the time of investment as in the setting in **Table 11**. The sample is a set of pairings of public pension investment entities with VC funds that were accessible at the time these pensions were making their investment decisions.

As in the analyses detailed in the previous section, I employ linear probability models to which I introduce variables indicative of public pensions' financial health - namely, the funded ratio and the discrepancy between assumed and actual rates

difference between assumed and realized rates on investments.

of return. These financial health metrics interact with the investment characteristics of past funds to assess how they might deferentially influence investment choices, depending on the financial status of the public pensions.

Table 13 reports the results. In Panel A, the funded ratio serves as the metric for assessing the financial health of public pensions. The results show that the funded ratio is not a significant determinant of public pension investment choices. However, the signs of the interaction term coefficients suggest that pensions that are well-funded are more inclined toward riskier investments. Panel B examines the impact of the return gap - the difference between assumed and realized returns. Here, significant results emerge: pensions facing a larger return gap are more inclined to select VC funds from firms with a history of less risky investments.

The results suggest a trend where underfunded pensions, rather than pursuing excessive risk, lean towards safer investment choices, potentially placing them in lower-yield, less risky VC funds. The subsequent section explores whether VC managers proactively consider the financial health of public pensions when shaping their investment portfolios.

1.4.7 Effect of Pension Financial Status on VC Investments

In this section, I examine whether the financial health of public pensions, either the funded ratio or the gap between assumed and realized returns, affects the decisions of VC managers. To that end, I focus on the investments made by VC funds backed by public pensions and use the same instrumental variable to instrument the aggregate financial health of pension investors in a VC fund. The overall financial well-being of pension investors is calculated as the weighted average of the funded ratio or return gap, with actuarial assets serving as the weighting factor. This aggregate financial health is instrumented as underfunded funds may opt for less risky VC funds, as discussed in the previous section.

Table 14 reports the results. Throughout the regressions, VC managers' investment choices do not appear to be markedly influenced by the financial well-being of their public pension investors. This contrasts sharply with the broader impact that the presence of public pensions has on VC investments. The lack of heterogeneity in this effect on VC investments suggests that VC managers might generalize public pensions, recognizing traits typically associated with such investors. These traits include stringent disclosure requirements, general financial objectives of the pensions, and concurrent trends that these pensions experience.

1.5 Conclusion

This paper shows that VC managers reduce the overall risks of their portfolios in the presence of public pensions. Specifically, they favor investments with a high likelihood of payoff, even if it means settling for lower returns. This approach results in a higher frequency of exits, including both acquisitions and IPOs, achieved in a shorter time frame.

To capture investment-level risk characteristics, I develop a novel measure of startups' technological links to public firms. This measure serves as an indicator of the startups' potential as attractive targets for public firms. Additionally, I utilize the number of patents and later-stage investments as tools through which VC managers might calibrate their portfolio risks. I find that VC managers choose startups that are closely linked to public firms, startups with more patents, and later-stage investments.

This finding is particularly fascinating given the VC industry's reputation for chasing a small number of highly lucrative deals. The study underscores that VC managers can actively mitigate risks by using investment-level risk metrics to fine-tune their portfolios' overall risk.

The adaptation of risk profiles to cater to the preferences of specific limited partners (LPs) might be distinctive to the VC industry. Unlike mutual fund managers, who do not typically emulate index funds to cater to public pensions as primary investors, VC managers might have incentives to take the characteristics of their investors into account given that the LP-General Partner (GP) dynamic often exhibits a deeper relationship. This connection is pivotal, as VC managers largely depend on existing investor relationships to secure funding for new funds. Furthermore, the significant size of certain investors within the VC sector, in contrast to other industries, provides additional incentive for VC managers to synchronize their investment strategies with investor preferences.

Public pensions do not appear to actively use these metrics to select VC funds. However, they seem to continue investing in the same VC funds if the funds they invested were managed in a less risky manner. This behavior aligns with VC managers' motivation to cater to their investors' preferences to maintain the relationship with them. I further find that rather than opting into more risky VC funds, underfunded public pensions tend to choose less risky ones. The financial health of public pensions does not significantly influence the decision-making of VC managers.

In conclusion, this study provides insights into the portfolio management strategies of VC managers who are incentivized to align with the interests of LPs. It demonstrates that VC managers have the necessary tools, namely investment-level risk metrics, to adjust risk levels according to the preferences of their LPs.

Table 1. Summary Statistics

The table presents the number of observations (N), mean, standard deviation (SD), minimum, 25th percentile (P25), median, 75th percentile (P75), and maximum of the variables in the paper. For start-up level variables, the sample contains startups funded by U.S. venture capital (VC) funds at least once between 1980 and 2017, but whose last investment was made before 2018. For VC funds, the sample contains US VC funds with a vintage year between 2001 and 2020. The variable $PublicTechBase\_Avg$ represents the proportion of inventions by public firms in the patent references cited by startups before their exits. This is averaged across all of a venture's inventions during its tenure as a startup. $log(RefIntensity)$ is the logarithm (plus one) of the number of patent references cited by the startup before an exit, normalized by the number of patent applications they submitted prior to that exit. $log(\#Patents)\_BeforeExit$ indicates the number of patents granted to a startup before its exit (incremented by 1), winsorized at 1 and 99 percentiles. In the absence of patent data for some startups, I derive the values for $PublicTechBase\_Avg$ and $log(RefIntensity)$ from similar startups possessing patent data, based on their business descriptions. These are denoted as $PublicTechBase\_Avg\_Imp$ and $log(RefIntensity)\_Imp$, respectively. $I(Exit)$ is an indicator variable that takes a value of one if a start-up successfully exits and zero otherwise. A successful exit refers to an exit through acquisitions or IPOs. $I(Exit\_Acq)$ is an indicator variable that takes a value of one if a start-up exits through an acquisition and zero otherwise. $I(Exit\_IPO)$ is an indicator variable that takes a value of one if a start-up exits through an IPO and zero otherwise. $PPFs$ is the number of public pension funds as LPs. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. For investment level variables, $I(Exit)$, $I(Exit\_Acq)$, and $I(Exit\_IPO)$ are indicator variables for a startup exits within fifteen years of the fund's inception. $PublicTechBase_Imp$ denotes the fraction of patents from public firms that are cited by a venture in its most recent patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other variables include the logarithm of the granted patents count as of the investment date (increased by one) ($log(\#Patents)$), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. For the variables for previous funds, $R(Exit)\_Past$ represents the exit ratio of startups that prior funds invested in, spanning three to six years before the focal fund's establishment. Exits are determined based on data preceding the fund's inception. Similarly, $R(Exit\_Acq)\_Past$ is the rate of exit through acquisitions, and $R(Exit\_IPO)\_Past$ is the rate of exit through IPOs. $PublicTechBase\_Past$ is the average value of $PublicTechBase\_Imp$ for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, $R(Early\_Inv)\_Past$ is the proportion of early-stage investments made by these prior funds during the same three- to six-year period.

Variable	N	Mean	SD	Min	P25	Median	P75	Max
Start-up level								
$PublicTechBase\_Avg\_Imp$	42379	0.225	0.144	0	0.129	0.198	0.298	1
$log(\#Patents)\_BeforeExit$	42376	0.519	1.029	0	0	0	0.693	4.443
$log(RefIntensity)\_Imp$	42379	2.731	0.712	2.079	2.398	2.767	3.115	6.964
$I(Exit)$	49084	0.411	0.492	0	0	0	1	1
$I(Exit\_Acq)$	49084	0.297	0.457	0	0	0	1	1
$I(Exit\_IPO)$	49084	0.115	0.319	0	0	0	0	1
Investment level								
$I(PPFs)$	129154	0.181	0.385	0	0	0	0	1
$I(PPFs\ in\ VC\ Firm)$	129154	0.285	0.451	0	0	0	1	1
$I(Exit)$	129154	0.294	0.456	0	0	0	1	1
$I(Exit\_Acq)$	129154	0.248	0.432	0	0	0	0	1
$I(Exit\_IPO)$	129154	0.046	0.209	0	0	0	0	1
Instrumental Variable	129154	0.211	0.408	0	0	0	0	1
$PublicTechBase\_Imp$	102005	0.231	0.165	0	0.122	0.2	0.319	1
$log(\#Patents)$	129154	0.223	0.625	0	0	0	0	3.258
Early Stage	129154	0.429	0.495	0	0	0	1	1
Recent Late Round Funding	129154	0.307	0.461	0	0	0	1	1
$log(VC\ Firm\ Age)$	128941	2.581	0.659	0	2.079	2.565	2.996	5.159
$Avg(log(\#Deals))$	129154	3.962	1.217	0.693	3.232	4.007	4.682	6.904
$Funded\ Ratio$	24819	3.962	0.196	0.196	0.028	0.899	0.995	1.476
$Return\ Gap$	129154	0.014	0.089	-0.094	-0.039	-0.000	0.035	0.571

Table 1. Summary Statistics - Continued

Variable	N	Mean	SD	Min	P25	Median	P75	Max
Investment with Exits								
Vintage to Exit	40717	8.623	4.080	0	5	8	11	22
VC fund level								
$I(PPFs)$	11257	0.077	0.267	0	0	0	0	1
Instrumental Variable	11257	0.171	0.377	0	0	0	0	1
Previous Funds' Records								
$PublicTechBase\_Past$	13078	0.125	0.141	0	0	0	0.242	1
$log(\#Patents)\_Past$	13078	0.216	0.394	0	0	0	0.313	4.304
$R(Early\_Inv)_Past$	13078	0.153	0.208	0	0	0	0.286	1
$R(Recent\_Late\_Round)\_Past$	13078	0.198	0.271	0	0	0	0.357	1
$I(Invested\ by\ Underfunded\ Pension)$	8532	0.101	0.302	0	0	0	0	1
$I(Invested\ by\ Low\ Return\ Pension)$	8532	0.054	0.227	0	0	0	0	1
Round-to-Round Returns: 1980-2022								
Round-to-Round Returns	35092	0.781	2.062	-1	-0.135	0.222	1.024	22.856
Pairs of Public Pension Investment Units and VC Funds as Accessible at the Time of Their Investments.								
Invested by a Public Pension	61390	0.033	0.179	0	0	0	0	1
$PublicTechBase\_Past$	59611	0.255	0.102	0	0.185	0.255	0.316	1
$log(\#Patents)\_Past$	61390	0.389	0.489	0	0	0.207	0.636	3.332
$R(Early\_Inv)_Past$	61390	0.403	0.179	0	0.176	0.347	0.599	1
$R(Recent\_Late\_Round)\_Past$	61390	0.288	0.229	0	0.083	0.284	0.433	1
$Avg(log(\#Deals))\_Past$	61390	3.858	1.275	0.693	3.044	3.850	4.672	7.374
$log(VC\ Firm\ Age)$	61390	2.759	0.616	0	2.302	2.708	3.178	5.153
$Past\ Client$	61390	0.023	0.149	0	0	0	0	1
$Funded\ Ratio$	39246	0.777	0.278	.0031	0.701	0.801	0.932	1.544
$Return\ Gap$	36180	0.057	0.193	-0.094	-0.031	-0.000	0.032	0.859

Table 2. Technology of Startups and Exit Likelihood

The table presents the results of OLS regressions that explore the relationship between the technology of a startup and exit likelihood. The dependent variable $I(Exit)$ is an indicator variable that takes a value of one if a startup exits either through acquisitions or IPOs, zero otherwise. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. In Panel A, the sample comprises startups with patent filings funded by venture capital firms from 1980 to 2017, with the last funding round occurring before 2018. Panel B includes startups with and without patent filings funded within the same period. The variable $PublicTechBase\_Avg$ represents the proportion of inventions by public firms in the patent references cited by startups before their exits. This is averaged across all of a venture's inventions during its tenure as a startup. $log(RefIntensity)$ is the logarithm (plus one) of the number of patent references cited by the startup before an exit, normalized by the number of patent applications they submitted prior to that exit. $log(\#Patents)\_BeforeExit$ indicates the number of patents granted to a startup before its exit. In the absence of patent data for some startups, I derive the values for $PublicTechBase\_Avg$ and $log(RefIntensity)$ from similar startups possessing patent data, based on their business descriptions. These are denoted as $PublicTechBase\_Avg\_Imp$ and $log(RefIntensity)\_Imp$, respectively. t-statistics, calculated using standard errors clustered at the intersection of the founding year and economic sector are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Panel A

	(1) $I(Exit)$	(2) $I(Exit\_Acq)$	(3) $I(Exit\_IPO)$	(4) $I(Exit)$	(5) $I(Exit\_Acq)$	(6) $I(Exit\_IPO)$
$PublicTechBase\_Avg$	0.256***	0.271***	-0.015	0.212***	0.172***	0.040**
	[6.18]	[8.13]	[-0.73]	[7.72]	[6.91]	[2.19]
$log(\#Patents)\_BeforeExit$	0.101***	0.023***	0.079***	0.092***	0.022***	0.070***
	[23.40]	[5.09]	[19.24]	[21.96]	[4.82]	[17.40]
$log(RefIntensity)\_Avg$	-0.001	0.020***	-0.021***	0.016***	0.030***	-0.014***
	[-0.21]	[4.03]	[-6.27]	[3.03]	[5.73]	[-4.65]
Sector×Founding Year FE				✓	✓	✓
Adjusted R^2	0.081	0.017	0.101	0.116	0.031	0.168
Observations	14157	14157	14157	13963	13963	13963

Panel B

	(1) $I(Exit)$	(2) $I(Exit\_Acq)$	(3) $I(Exit\_IPO)$	(4) $I(Exit)$	(5) $I(Exit\_Acq)$	(6) $I(Exit\_IPO)$
$PublicTechBase\_Avg\_Imp$	0.207***	0.284***	-0.077***	0.194***	0.186***	0.008
	[5.06]	[10.24]	[-3.11]	[10.44]	[9.87]	[0.68]
$log(\#Patents)\_BeforeExit$	0.063***	0.027***	0.035***	0.057***	0.028***	0.029***
	[16.89]	[8.80]	[12.64]	[24.41]	[9.55]	[13.41]
$log(RefIntensity)\_Imp$	0.019***	0.013***	0.006**	0.027***	0.017***	0.010***
	[4.86]	[3.48]	[2.56]	[7.60]	[4.82]	[4.41]
Sector×Founding Year FE				✓	✓	✓
Adjusted R^2	0.025	0.013	0.014	0.079	0.023	0.113
Observations	42376	42376	42376	42175	42175	42175

Table 3. Public Pension Investors and Deal Characteristics: Linear Relationship

The table presents the linear relationships between the presence of public pension funds as LPs and investment attributes related to risk. The sample consists of investments made by U.S. VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. The variable $PublicTechBase_Imp$ denotes the fraction of patents from public firms that are cited by a venture in its most recent patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. *Avg*(*log*(#*Deals*)) is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	PublicTechBase_Imp			*log*(#*Patents*)		
	(1)	(2)	(3)	(4)	(5)	(6)
$I(PPFs)$	0.022***	0.015***	0.015***	0.094***	0.058***	0.029
	[4.02]	[3.24]	[3.22]	[4.98]	[2.96]	[1.57]
log(*VC Firm Age*)			-0.002			0.077***
			[-0.71]			[7.39]
Avg(*log*(#*Deals*))			0.001			-0.030***
			[0.57]			[-5.38]
Fund Focus×Fund Founding Year FE		✓	✓		✓	✓
Fund Stage FE		✓	✓		✓	✓
Adjusted R^2	0.003	0.049	0.049	0.003	0.037	0.043
Observations	102005	101999	101827	129154	129147	128934
	Early Stage			Recent Late Round Funding		
	(7)	(8)	(9)	(10)	(11)	(12)
$I(PPFs)$	-0.179***	-0.106***	-0.076***	0.173***	0.115***	0.081***
	[-11.06]	[-8.32]	[-6.90]	[12.43]	[9.07]	[7.13]
log(*VC Firm Age*)			-0.091***			0.085***
			[-13.92]			[11.08]
Avg(*log*(#*Deals*))			0.048***			-0.026***
			[6.14]			[-3.64]
Fund Focus×Fund Founding Year FE		✓	✓		✓	✓
Fund Stage FE		✓	✓		✓	✓
Adjusted R^2	0.019	0.104	0.12	0.021	0.069	0.08
Observations	129154	129147	128934	129154	129147	128934

Table 4. Test of Relevance Condition

The table shows that the instrumental variable correlates with the presence of public pensions as investors. In Panel A, the sample contains U.S. VC funds with a vintage year between 2001 and 2021. The dependent variable is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. The instrumental variable is an indicator variable that takes a value of one if a VC fund was established due to the typical VC fundraising cycle, and its creation was in a state experiencing a favorable chance of receiving funding from public pensions as of the capital raising phase. In Panel B, the same variables are employed, but the sample comprises investments made by VC funds. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Panel A

	(1)	(2)	(3)
Instrumental Variable	0.132***	0.130***	0.125***
	[9.77]	[10.89]	[10.69]
Fund Focus×Fund Founding Year FE		✓	✓
Fund Stage FE			✓
Adjusted R^2	0.037	0.062	0.08
Observations	11274	11272	11265

Panel B

	(1)	(2)	(3)
Instrumental Variable	0.225***	0.174***	0.146***
	[6.21]	[5.95]	[7.07]
Log(VC Firm Age)		0.106***	
		[6.65]	
Avg(Log(#Deals))		0.004	
		[0.67]	
Fund Founding Year FE		✓	
Fund Stage FE		✓	
Fund Focus×Fund Founding Year ×Fund Stage×Age Group×Size Group			✓
Adjusted R^2	0.061	0.196	0.488
Observations	154060	153810	153642

Table 5. Effect of Public Pension Funds on Deal Characteristics

The table shows the effect of public pension funds on investment attributes related to risk, estimated from the second-stage regressions. The sample contains investments made by US VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a comprehensive set of indicator variables to achieve full saturation of control variables. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a comprehensive set of indicator variables to achieve full saturation of control variables. The variable $PublicTechBase_Imp$ denotes the fraction of patents from public firms that are cited by a venture in its most recent patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$			$log(\#Patents)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$I(\widehat{PPFs})$	0.059***	0.032***	0.025**	0.326***	0.203***	0.168***
	[3.29]	[2.91]	[2.16]	[6.77]	[3.87]	[3.16]
Fund Focus	✓	✓	✓	✓	✓	✓
Fund Founding Year	✓	✓	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓	✓	✓
Age Group		✓	✓		✓	✓
Size Group			✓			✓
Adjusted R^2	0.058	0.075	0.086	0.053	0.061	0.075
Observations	101911	101787	101673	129067	128940	128823
	Early Stage			Recent Late Round Funding		
	(7)	(8)	(9)	(10)	(11)	(12)
$I(\widehat{PPFs})$	-0.248***	-0.154***	-0.107***	0.293***	0.191***	0.144***
	[-5.75]	[-3.55]	[-3.79]	[7.32]	[4.42]	[5.25]
Fund Focus	✓	✓	✓	✓	✓	✓
Fund Founding Year	✓	✓	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓	✓	✓
Age Group		✓	✓		✓	✓
Size Group			✓			✓
Adjusted R^2	0.12	0.134	0.155	0.085	0.097	0.116
Observations	129067	128940	128823	129067	128940	128823

Table 6. Effect of Public Pension Funds on Startup Exits

The table shows the effect of public pension funds on the likelihood of exits, estimated from the second-stage regressions. The sample contains investments made by U.S. VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its investor and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables for combinations of the values of the control variables. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a comprehensive set of indicator variables to achieve full saturation of control variables. $I(Exit)$ is an indicator that assumes a value of one if a startup exits - either through acquisitions or IPOs - within fifteen years of the fund's inception. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$I(Exit\_Acq)$			$I(Exit\_IPO)$			$I(Exit)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$I(\widehat{PPFs})$	0.101***	0.128***	0.074***	0.118***	0.088***	0.070***	0.219***	0.216***	0.144***
	[3.52]	[3.75]	[2.89]	[5.97]	[5.71]	[4.68]	[6.31]	[5.74]	[5.77]
Fund Focus	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fund Founding Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age Group		✓	✓		✓	✓		✓	✓
Size Group			✓			✓			✓
Adjusted R^2	0.037	0.067	0.082	0.041	0.059	0.074	0.046	0.077	0.094
Observations	129067	128940	128823	129067	128940	128823	129067	128940	128823

Table 7. Effect of Public Pension Funds on the Time to Exit in VC Investments

The table presents the impact of public pension funds on investment exit timing, as estimated from the second-stage regression analyses. The sample comprises investments made by U.S.-based VC funds established between 2001 and 2017 that ultimately achieved successful exits. The dependent variable is the duration, measured in years, between the founding year of the VC fund and the occurrence of a successful investment exit, limited to exits that take place within 15 years of the fund's inception. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a comprehensive set of indicator variables to achieve full saturation of control variables. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables for combinations of the values of control variables. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

All exits	(1)	(2)	(3)
$I(\widehat{PPFs})$	-2.334***	-2.021***	-1.658***
	[-6.16]	[-5.15]	[-5.70]
Fund Focus	✓	✓	✓
Fund Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.292	0.314	0.35
Observations	37891	37794	37696

Acquisitions	(1)	(2)	(3)
$I(\widehat{PPFs})$	-2.274***	-1.904***	-1.597***
	[-6.26]	[-4.31]	[-4.78]
Fund Focus	✓	✓	✓
Fund Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.315	0.336	0.372
Observations	31987	31904	31811

IPOs	(1)	(2)	(3)
$I(\widehat{PPFs})$	-1.543**	-1.418***	-1.197**
	[-2.26]	[-3.68]	[-2.20]
Fund Focus	✓	✓	✓
Fund Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.265	0.294	0.354
Observations	5841	5781	5719

Table 8. Presence of Public Pensions in a VC Firm and Investment Characteristics

The table shows the relationship between the presence of public pensions and VC firm-wide investments. The dependent variables are investment-level characteristics. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms that are cited by a venture in its most recent patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator variable for startups that have secured late-round capital within the preceding two years. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. $I(PPFs\ in\ VC\ Firm)$ is an indicator variable that takes the value of one if a VC fund, which has public pension investors, was created within the same VC firm in the past 10 years. Otherwise, it takes a value of zero. $I(PPFs)$ is an indicator variable that takes one if there is a public pension as an LP, zero otherwise. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$		$log(\#Patents)$		Early Stage		Recent Late Round Funding	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$I(PPFs\ in\ VC\ Firm)$	-0.001	-0.003	-0.070**	-0.078***	0.024	0.034*	-0.024	-0.033
	[-0.24]	[-0.63]	[-2.55]	[-2.66]	[1.58]	[1.96]	[-1.15]	[-1.47]
$I(PPFs)$		0.006*		0.028*		-0.034***		0.032***
		[1.72]		[1.76]		[-2.93]		[2.88]
$log(VC\ Firm\ Age)$	-0.048***	-0.048***	0.190***	0.190***	-0.165***	-0.165***	0.153***	0.153***
	[-13.19]	[-13.18]	[9.19]	[9.26]	[-12.28]	[-12.43]	[9.19]	[9.30]
$Avg(log(\#Deals))$	-0.002	-0.002	-0.016*	-0.016*	0.027**	0.026**	-0.026**	-0.025**
	[-0.61]	[-0.56]	[-1.77]	[-1.69]	[2.11]	[2.06]	[-2.29]	[-2.23]
VC Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Focus×Fund Founding Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.147	0.147	0.115	0.115	0.217	0.217	0.165	0.165
Observations	112608	112608	145625	145625	145625	145625	145625	145625

Table 9. Investment Characteristics and Round-to-Round Returns

The table presents the relationship between investment characteristics and round-to-round returns. The sample includes investments by U.S.-based VC funds from 1980 to 2022 with available round-to-round returns. The dependent variable represents returns between the current funding round and the subsequent one. The variable $PublicTechBase\_Imp$ denotes the fraction of patents from public firms that are cited by a venture in its most recent patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other investment characteristics include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator variable for late-round investments that have secured late-round capital in the preceding two years. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)
$PublicTechBase\_Imp$	-0.031			
	[-0.46]			
$log(\#Patents)$		-0.131***		
		[-5.75]		
$Early\ Stage$			0.505***	
			[8.76]	
$Recent\ Late\ Round\ Funding$				-0.443***
				[-10.02]
$Avg(log(\#Deals))$	0.087***	0.071***	0.068***	0.066***
	[5.27]	[3.34]	[3.20]	[3.11]
$log(VC\ Firm\ Age)$	-0.054*	-0.094*	-0.077	-0.072
	[-1.93]	[-1.88]	[-1.56]	[-1.43]
Fund Focus×Fund Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.017	0.006	0.01	0.009
Observations	24858	31466	31466	31466

Table 10. Previous Funds' Characteristics and the Presence of Public Pensions

The table shows which characteristics of previous funds explain the presence of public pension funds in a VC fund. It comprises VC funds established between 2001 and 2020, which had a record of investments of earlier VC funds three to six years prior to their inception, inclusively. The dependent variable is an indicator variable that takes a value of one if the fund has a public pension as an investor and zero otherwise. *PublicTechBase_Past* is the average value of *PublicTechBase_Imp* for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, *log(#Patents)_Past* represents the mean of the natural logarithm of the total number of patents (with an addition of one) from past investments. *R(Early_Inv)_Past* denotes the share of early-stage investments conducted by these previous funds during the same three- to six-year timeframe. Likewise, *R(Recent_Late_Round)_Past* indicates the proportion of late-round investments with a history of another late round within the past two years executed by these previous funds. *Avg(log(#Deals))* is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of fund focus, fund stage, and founding year levels, are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)
PublicTechBase_Past	0.322***			
	[8.81]			
log(#Patents)_Past		0.057***		
		[4.09]		
R(Early_Inv)_Past			0.003	
			[0.20]	
R(Recent_Late_Round)_Past				0.256***
				[9.87]
Avg(log(#Deals))	0.023***	0.030***	0.033***	0.026***
	[8.88]	[11.76]	[12.87]	[10.38]
log(VC Firm Age)	0.005	0.019***	0.029***	0.001
	[0.94]	[3.55]	[5.41]	[0.15]
Fund Focus×Fund Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.117	0.11	0.106	0.124
Observations	11690	11690	11690	11690

Table 11. Previous Funds' Characteristics and Public Pensions' Decisions

The table shows which characteristics of prior funds explain the investment decisions of individual public pension funds, based on linear probability models. The sample includes distinct combinations of public pension investment units and VC funds available for investment when these pensions made their decisions. A VC fund is deemed accessible to public pension investors if it was established in the same year, the year before, or the year after the public pension made an investment in any VC fund, and it was located within the five states closest to the public pension. I require that VC funds are created between 2001 and 2020 and have investment records from earlier VC funds dating back three to six years prior to their establishment. The dependent variable is an indicator that takes a value of one if a fund was chosen by a public pension investment unit, and zero otherwise. *Past Client* is an indicator variable that takes one if a public pension investment unit was a client of the VC firm of the VC fund before, zero otherwise. *PublicTechBase_Past* is the average value of *PublicTechBase_Imp* for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, *log(#Patents)_Past* represents the mean of the natural logarithm of the total number of patents (with an addition of one) from past investments. *R(Early_Inv)_Past* denotes the share of early-stage investments conducted by these previous funds during the same three- to six-year timeframe. Likewise, *R(Recent_Late_Round)_Past* indicates the proportion of late-round investments with a history of another late round within the past two years executed by these previous funds. *Avg(log(#Deals))_Past* is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of public pension investment unit, public pension investment year, and fund focus are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PublicTechBase_Past	-0.030***	-0.015***						
	[-4.02]	[-4.24]						
*PublicTechBase_Past*Past Client*		0.505***						
		[6.20]						
log(#Patents)_Past			0.007***	0.003***				
			[3.69]	[3.69]				
*log(#Patents)_Past*Past Client*				0.044***				
				[5.05]				
R(Early_Inv)_Past					-0.014***	-0.001		
					[-4.75]	[-1.01]		
*R(Early_Inv)_Past*Past Client*						-0.129***		
						[-5.57]		
R(Recent_Late_Round)_Past							0.026***	0.007***
							[6.20]	[3.40]
*R(Recent_Late_Round)_Past*Past Client*								0.059***
								[3.08]
Past Client		0.801***		0.913***		0.972***		0.911***
		[32.55]		[117.63]		[184.09]		[90.97]
Avg(log(#Deals))_Past	0.007***	0.002***	0.007***	0.001***	0.007***	0.002***	0.007***	0.002***
	[8.64]	[4.20]	[8.40]	[3.88]	[8.69]	[4.23]	[8.74]	[4.29]
log(VC Firm Age)	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001	-0.003*	-0.001
	[-0.30]	[-0.81]	[-0.71]	[-1.26]	[-0.94]	[-0.68]	[-1.71]	[-1.41]
Pension Investment Unit× Pension Investment Year× Fund Focus FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.112	0.69	0.113	0.69	0.113	0.691	0.113	0.69
Observations	57473	57473	59028	59028	59028	59028	59028	59028

Table 12. Past VC Fund Characteristics and the Presence of Underfunded Pensions

The table shows which characteristics of previous funds are considered by public pension funds based on their funded status and past investment returns. The dataset includes public pension-backed VC funds established between 2001 and 2021. These funds had a history of investments from previous VC funds spanning three to six years before their creation. In the first four columns, the dependent variable is an indicator variable that takes a value of one if the fund has a public pension as an investor whose funded ratio is below 0.9, and zero otherwise. In the subsequent columns, the dependent variable is set to one if the fund has a public pension investor with a difference between its Assumed Ratio of Returns (ARR) and the investment return of the past three years exceeding 0.03. Otherwise, it's set to zero. $PublicTechBase\_Past$ is the average value of $PublicTechBase\_Imp$ for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, $log(\#Patents)\_Past$ represents the average of the natural logarithm of the total number of patents (with an addition of one) from past investments. $R(Early\_Inv)_Past$ denotes the share of early-stage investments conducted by these previous funds during the same three- to six-year timeframe. Likewise, $R(Recent\_Late\_Round)\_Past$ indicates the proportion of late-round investments with a history of another late round within the past two years executed by these previous funds. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Invested by Underfunded Pensions				Invested by Pensions Not Keeping Up with ARR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$PublicTechBase\_Past$	0.394**				0.033			
	[2.48]				[0.22]			
$log(\#Patents)\_Past$		0.001				0.035		
		[0.02]				[0.85]		
$R(Early\_Inv)_Past$			-0.1				-0.135**	
			[-1.19]				[-2.47]	
$R(Recent\_Late\_Round)\_Past$				0.024				0.014
				[0.26]				[0.15]
$Avg(log(\#Deals))$	0.089***	0.096***	0.101***	0.096***	0.067***	0.065***	0.074***	0.067***
	[4.41]	[5.13]	[4.90]	[4.89]	[2.70]	[2.75]	[3.02]	[2.81]
$log(VC\ Firm\ Age)$	-0.090***	-0.062*	-0.062*	-0.065*	-0.01	-0.016	-0.008	-0.01
	[-2.67]	[-1.74]	[-1.86]	[-1.91]	[-0.43]	[-0.70]	[-0.40]	[-0.35]
Fund Focus×Fund Founding Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.271	0.266	0.267	0.266	0.489	0.489	0.491	0.489
Observations	1171	1171	1171	1171	1171	1171	1171	1171

Table 13. Previous Funds' Characteristics and Public Pensions' Financial Status

The table indicates which characteristics of prior funds explain the investment decisions of individual public pension funds, based on linear probability models. The sample includes distinct combinations of public pension investment units and VC funds available for investment when these pensions made their decisions. A A VC fund is deemed accessible to public pension investors if it was established in the same year, the year before, or the year after the public pension made an investment in any VC fund, and it was located within the five states closest to the public pension. I require that VC funds are created between 2001 and 2020 and have investment records from earlier VC funds dating back three to six years prior to their establishment. The dependent variable is an indicator that takes a value of one if a fund was chosen by a public pension investment unit, and zero otherwise. *PublicTechBase_Past* is the average value of *PublicTechBase_Imp* for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, *log(#Patents)_Past* represents the mean of the natural logarithm of the total number of patents (with an addition of one) from past investments. *R(Early_Inv)_Past* denotes the share of early-stage investments conducted by these previous funds during the same three- to six-year timeframe. Likewise, *R(Recent_Late_Round)_Past* indicates the proportion of late-round investments with a history of another late round within the past two years executed by these previous funds. *Avg(log(#Deals))_Past* is the average number of investments made by previous funds. In Panel A, I investigate the relationship between the funded ratio, weighted-averaged across public pensions in a public pension investment unit, using total assets as weights (*Funded Ratio*) and the fund choice of the investment unit. In Panel B, I examine the relationship between the gap between assumed and realized past three-year investment returns, weighted-averaged across public pensions in a pension investment unit (*Return Gap*), and the fund choice of the investment unit. t-statistics, calculated using standard errors clustered at the intersection of public pension investment unit, public pension investment year, and fund focus levels are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Panel A			
	(1)	(2)	(3)	(4)
PublicTechBase_Past	-0.003			
	[-0.11]			
*PublicTechBase_Past*Funded Ratio*	-0.038			
	[-1.28]			
log(#Patents)_Past		0.015**		
		[2.56]		
*log(#Patents)_Past*Funded Ratio*		-0.015**		
		[-2.10]		
R(Early_Inv)_Past			-0.021**	
			[-2.32]	
*R(Early_Inv)_Past*Funded Ratio*			0.012	
			[1.03]	
R(Recent_Late_Round)_Past				0.038***
				[2.97]
*R(Recent_Late_Round)_Past*Funded Ratio*				-0.019
				[-1.17]
Funded; Ratio	0.008	0	-0.007	0.001
	[0.33]	[-0.02]	[-0.33]	[0.06]
Avg(log(#Deals))_Past	0.007***	0.007***	0.007***	0.007***
	[6.73]	[6.50]	[6.75]	[6.82]
log(VC Firm Age)	-0.007***	-0.007***	-0.008***	-0.009***
	[-3.13]	[-3.09]	[-3.38]	[-3.90]
Pension Investment Unit FE	✓	✓	✓	✓
Pension Investment Year Fund Focus FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.032	0.032	0.032	0.033
Observations	37880	38889	38889	38889

Table 13. Previous Funds' Characteristics and Public Pensions' Financials - Continued

Panel B				
	(1)	(2)	(3)	(4)
PublicTechBase_Past	-0.040***			
	[-3.82]			
*PublicTechBase_Past*Return Gap*	0.127***			
	[3.06]			
log(#Patents)_Past		0.003		
		[1.21]		
*log(#Patents)_Past*Return Gap*		0.004		
		[0.39]		
R(Early_Inv)_Past			-0.009**	
			[-2.10]	
*R(Early_Inv)_Past*Return Gap*			-0.046***	
			[-2.75]	
R(Recent_Late_Round)_Past				0.022***
				[3.94]
*R(Recent_Late_Round)_Past*Return Gap*				0.057**
				[2.39]
Return Gap	-0.158***	-0.108***	-0.116***	-0.143***
	[-4.10]	[-3.10]	[-3.31]	[-3.67]
Avg(log(#Deals))_Past	0.008***	0.008***	0.008***	0.008***
	[7.08]	[6.85]	[7.08]	[7.21]
log(VC Firm Age)	-0.008***	-0.008***	-0.008***	-0.010***
	[-3.41]	[-3.28]	[-3.64]	[-4.29]
Pension Investment Unit FE	✓	✓	✓	✓
Pension Investment Year Fund Focus FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.035	0.034	0.035	0.035
Observations	34967	35876	35876	35876

Table 14. Effect of Pension Financial Status on VC Investments

The table shows the effect of the financial status of public pension investors on VC managers' investment decisions. The sample is comprised of investments made by U.S. VC funds with public pension investors established between 2001 and 2017. In the first four regressions, $\widehat{Funded\ Ratio}$ is the predicted value obtained from the first-stage regression. In the first-stage regression, the dependent variable is the weighted average funded ratio, where the weights are determined by the total assets of public pensions. For the subsequent regressions, $\widehat{Return\ Gap}$ is the predicted value obtained from the first-stage regression. In the first-stage regression, the dependent variable is the weighted average difference between the assumed rate of returns and the investment returns from the past three years. The weights are based on the total assets of the public pensions. The first-stage regressions in all columns incorporate indicator variables that represent combinations of values for the instrumental variable and all control variables. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for late-round investments that have secured late-round funding within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
	(1)	(2)	(3)	(4)
$\widehat{Funded\ Ratio}$	-0.027*	0.082	-0.006	0.033
	[-1.65]	[1.04]	[-0.20]	[0.78]
Fund Focus	✓	✓	✓	✓
Fund Founding Year	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓
Adjusted R^2	0.068	0.054	0.131	0.064
Observations	18461	22910	22910	22910
	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
	(5)	(6)	(7)	(8)
$\widehat{Return\ Gap}$	0.05	-0.157	0.12	-0.06
	[1.22]	[-1.02]	[1.14]	[-0.64]
Fund Focus	✓	✓	✓	✓
Fund Founding Year	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓
Adjusted R^2	0.061	0.054	0.133	0.063
Observations	17823	22078	22078	22078

Appendix A: Business Description-Based Company Relationship Prediction

In this section, I detail the process involved in developing and training a deep-learning model. The primary purpose of this model is to estimate the business similarity (or the extent of technological overlap) between two companies, using the data derived from their business descriptions.

A.1. Link Prediction in Deep Learning

Link prediction in deep learning involves using neural network models to forecast the probability of a connection or relationship forming between two entities within a network. Link prediction models learn from the network's structure and relevant features to estimate the likelihood of missing connections based on given data, which is valuable in this research where it aids in uncovering a hidden link between two startups.

A.2. Understanding the Role of the Model in the Paper

The paper employs a metric designed to capture the technological relationships of startups with public firms. This metric relies on patent information, yet it is noteworthy that only approximately one-third of these startups possess patent application data (see **Table A.1**). To overcome this issue, I infer the technological orientation of startups lacking patent data from those with available patent data. To identify suitable matches for startups lacking patent information, I need a matching method. However, VentureXpert does not provide the kind of structured information, like categories or numerical values, that would make it easy to find matching startups. Consequently, the study leverages a richer source of information, business descriptions provided by the database. I create a deep learning model adept at comprehending intricate textual information to identify a business relatedness between two given startups.

A.3. Model Architecture

The model processes business descriptions from two companies to generate a link probability that connects these firms. In this section, I will outline the key aspects of the deep learning model, while a more detailed description of the architecture can be found in **Figure A.1 and Figure A.2**.

A.3.1. Inputs and Outputs

The model takes the embeddings of business keywords from two companies, without considering any specific order. While a company can have multiple business keywords, due to resource constraints, I limit the maximum number of keywords to thirty. The output of the model is a number between zero to one that will be used to determine if two companies are meaningfully related in terms of business or technologies.

A.3.2. Asymmetric Association Scores

The key innovation of the model is that it employs a layer that captures the asymmetric dependency between two entities. Although similarity scores measured by cosign similarities between any two keywords of two startups can show the level of connection between two startups, there may be instances where a single business keyword from one startup is connected to many keywords from another. In such cases, the overall similarity score might be low. To overcome this, a layer is required that can amplify the scores when even a few keywords are related to those of another firm. This ensures that connections are not undervalued due to the presence of unrelated keywords.

To achieve the goal, I create a layer of the model that computes Asymmetric Association Scores (A→B) ($AAS_{A\to B}$) between an entity A and another entity B, representing A's relatedness to B.

Let's denote a matrix, A, which contains rows representing each keyword of the first company. Similarly, for the second company, this matrix is denoted as B. $AAS_{A\to B}$ in a particular layer i is computed as follows.

$$
\begin{aligned}
AAS_{i,A\to B} =& \tilde{A}_i \left(softmax_{dim=1} \left(\frac{\hat{A}_i \hat{B}_i^{\,T}}{\sqrt{d}} \right) \tilde{B}_i \right)^T \frac{1}{\sqrt{d}} \\
\tilde{A}_i &= AW_i^{Aa} \\
\tilde{B}_i &= BW_i^{Ba} \\
\hat{A}_i &= AW_i^{Ab} \\
\hat{B}_i &= BW_i^{Bb}
\end{aligned}
$$

W_i^{Aa}, W_i^{Ba}, W_i^{Ab}, and W_i^{Bb} are weights of the same dimension to be trained. d is the dimension of the weights. Scaling the matrix multiplication by the square root of the dimension size helps in generating stable gradients. The function $softmax_{dim=1}$

normalizes each row of the input matrix so that the sum of elements in any row equals one.

The formula indicates that if a good portion of keywords from the first company (whose matrix is A) is similar to even just a few keywords from the second company (whose matrix is B), the asymmetric association scores will be high overall. However, this effect does not hold when only a few keywords of the first company are similar to a large portion of the keywords of the second company. To account for this opposite scenario and to make the model order-agnostic, I also incorporate $AAS_{i,B \to A}$ into the model.

A.3.3. Preparing Inputs

The input for the model (keywords) is extracted from business descriptions, with a primary focus on the products and services that companies offer.

First, I extract keywords from the long business description provided by the database using a keyword extractor YAKE to avoid breaching data policy. YAKE is a lightweight, unsupervised keyword extraction method. It relies on statistical features derived from individual documents to identify the most significant keywords within a text.[12] I filter the keywords using Named Entity Recognition by Spacy to remove the names of companies.

Then, I input the keywords to a large language model to find more nouns related to products/services/technologies of a firm from OpenAI ChatGPT 3.5 Turbo. In this step, I ask the LLM to provide items related to the keywords, focusing on products, services, and technologies. The LLM is instructed to provide items related to the keywords without offering additional explanations or context.

After I find a set of unique keywords from the LLM, I extract embeddings of the keywords from another large language model, Open AI's ada-002. Embeddings from the model offer valuable features for natural language processing tasks. They capture both semantic and syntactic relationships between words in a context-dependent manner, allowing a nuanced understanding of language.

[12] Relying on this feature, I use a keywords extractor to isolate the 10 keywords that have the maximum N-gram size of seven from the business description. If I increase the number of keywords to more than 10, the extractor starts to produce keywords highly overlapping one another as the length of the long business description is typically one paragraph. Also, if the length of the text is short, then extracted keywords include terms that are not related to offerings/technologies/science of startups. I append the final section of the long business description, which lists the keywords provided by the database. To enrich the information, I supplement this with more offerings/technologies/science by querying the OpenAI's ChatGPT 3.5 Turbo.

A.3.4. Preparing Labels

The objective of the model is to identify a connection between two startups, necessitating a dataset of labels that depict the relationships between them. My training and test datasets focus on startups with patent data, as patent data offers valuable information to infer the relationships between these startups. To be included in the training and test sets, startups must have complete information on their founding date, exit date (either through successful exits or bankruptcy), at least one patent publication released between the founding and exit dates, and a business description available in VentureXpert.

To generate a set of labels, I determine the relationship between two startups by analyzing their patent citations while they were operating as startups. A link is created when one startup cites another's patents, I consider all patents within two hops in the citation network. For example, if startup A cites or is cited by a patent which, in turn, cites or is cited by startup B's patents — regardless of whether the intermediary patent belongs to a private or public company — they are still considered to be connected. In the sample, I find 5,428,756 links among startups with patent data. 17,548 are unique startups with patents that have at least one relationship with other startups.

The label is set to one if two startups are connected by a citation network within two hops. If not, the label is assigned a value of zero.

A.3.4. Training and Test Sets

I construct a balanced training set comprising an equal number of connected and unconnected startup pairs, specifically selecting 100,000 pairs each of connected and disconnected startups. For the test set, I select 20,000 pairs of startups not included in the training set, evenly split between 10,000 connected and 10,000 disconnected pairs.

A.3.5. Performance Evaluation on the Test Set

A.4. Finding Startup Matches with Patent Data

For each of the startups lacking patent data, I feed all possible pairs of the startup and a startup from the pool of startups with patent data into the finalized model. I apply a threshold of 0.6, as opposed to 0.5, to enhance the precision of the output. Precision in a binary classification model refers to the proportion of true positive predictions (startups correctly identified as having a significant relationship) among all positive predictions. By setting a higher threshold, I aim to reduce the number of false positives (startups incorrectly identified as having a significant relationship), thereby increasing precision.

A.5. Imputed Value of $PublicTechBase$ for Non-Patent startups

For startups that have filed patent applications, I can find the precise year of their patent application submissions. I leverage this data to compute time-varying tech characteristics, denoted as $PublicTechBase$, which quantify the degree to which a startup's technology aligns with that of public firms. To achieve this, I create a venture-year dataset and apply forward filling followed by backward filling. For startups that have not submitted patent applications, I calculate their imputed tech characteristics using the average $PublicTechBase$ value from matched startups that possess patent data for the corresponding year.

Table A.1: Share of Startups with Patent Applications and Those Citing Public Firms' Patents

The table shows the number of total startups, number of startups with patent applications, share of startups with patent applications, number of startups that cite public firms' patents, and share of startups that cite public firms' patents. The sample contains startups backed by US VC funds from 1980 to 2022.

	Total	Technology	Healthcare	Industrials	Consumer Cyclicals
Total number of firms (A)	72471	38400	11139	7569	6057
Number of startups with patent applications (B)	23213	11728	5912	2494	1206
Share of startups with patent applications (B)/(A)	32.03%	30.54%	53.07%	32.95%	19.91%
Number of startups citing public firms' patents (C)	17910	9173	4719	1866	844
Share of startups citing public firms' patents (C)/(B)	77.16%	78.21%	79.82%	74.82%	69.98%
	Consumer Non-Cyclicals	Financials	Basic Materials	Energy	Other
Total number of firms (A)	3083	2573	1423	851	1376
Number of startups with patent applications (B)	523	255	642	284	169
Share of startups with patent applications (B)/(A)	16.96%	9.91%	45.12%	33.37%	12.28%
Number of startups citing public firms' patents (C)	341	167	487	210	103
Share of startups citing public firms' patents (C)/(B)	65.20%	65.49%	75.86%	73.94%	60.95%

Figure A.1 Company Link Prediction Model Architecture

This figure depicts the structure of the deep learning model, which is designed to identify the technological connection between two startups. ❶ The boxes labeled A and B represent the sets of business keywords for startups A and B, respectively. These keywords are converted into embeddings of 1536 numbers using OpenAI's Ada 002. ❷ A linear dense layer, utilizing Exponential Linear Unit (ELU) activation functions, is employed. Each keyword retains a feature dimensionality of 1536. The linear layers enclosed within the dashed-line boxes possess identical properties. ❸ The connection layer, which is detailed below, computes the asymmetric association scores and similarity scores of the embeddings of two startups after a linear transformation. It then aggregates this information. ❹ An ELU-activated linear dense layer is employed. It takes an input of 90 dimensions and transforms it into an output of 82 dimensions. ❺ A linear dense layer, equipped with ELU activation functions, takes in an input of 164 dimensions and produces an output vector consisting of 82 elements. ❻ The output layer is composed of a dense layer that transitions from 82*3 to 82 dimensions, followed by an ELU activation function. Subsequently, the tensor is flattened and passed into another dense layer, which outputs a single number. This number is then processed through a sigmoid function. The final output is a numerical value ranging from zero to one, indicating the degree of interconnectedness between the two startups.

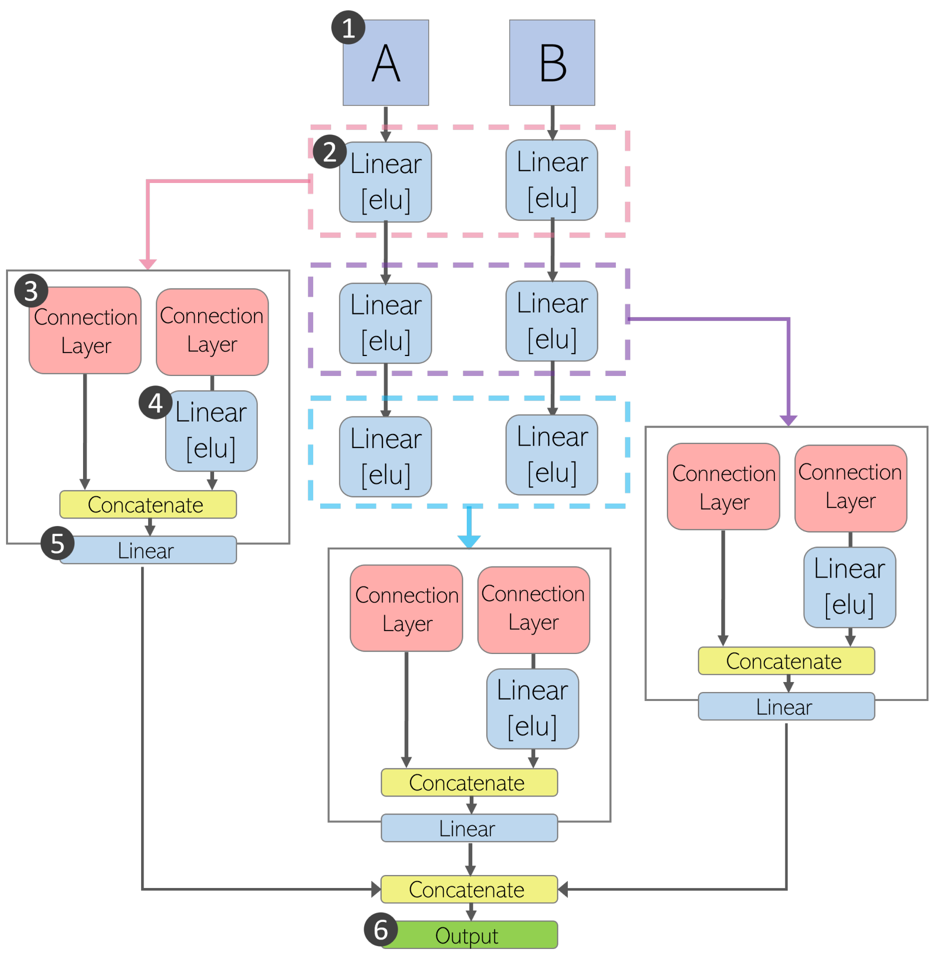
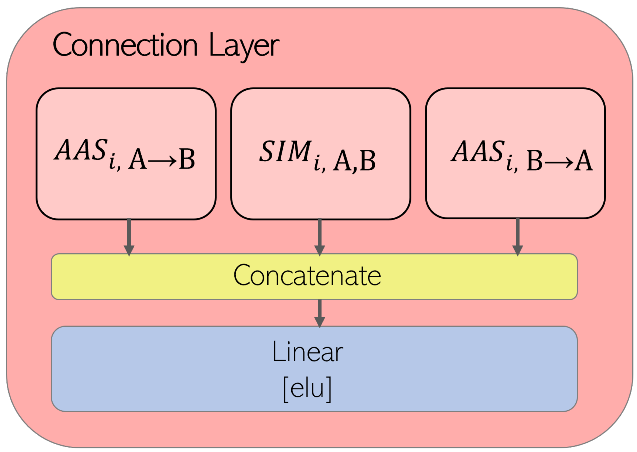


Figure A.2 Connection Layer

This figure illustrates the architecture of the connection layer. Two sets of embeddings from a previous layer denoted by A and B, each representing two startups, are input into the functions $AAS_{i,A\rightarrow B}$, $SIM_{i,A,B}$, and $AAS_{i,B\rightarrow A}$. Each function produces a 30 by 30 matrix, corresponding to the 30 keywords. The matrices denoted by W with varying superscripts and subscripts represent trainable weights. d stands for the feature dimension. The outputs of these functions are then concatenated and passed through a linear layer, which yields a matrix of 82 dimensions. This is subsequently processed by an ELU activation function.



Asymmetric Association Score Layer

$$AAS_{i,\,\mathrm{A}\rightarrow\mathrm{B}} = \tilde{A}_i \left(softmax_{dim=0}\left(\frac{\widehat{A}_i \widehat{B}_i^{T}}{\sqrt{d}} \right) \tilde{B}_i \right)^{T} \frac{1}{\sqrt{d}}$$

$$\tilde{A}_i = W_i^{Aa} A \qquad \widehat{A}_i = W_i^{Ab} A$$

$$\tilde{B}_i = W_i^{Ba} B \qquad \widehat{B}_i = W_i^{Bb} B$$

Similarity Score Layer

$$SIM_{i,\,\mathrm{A,B}} = cosine\ similarity(\overline{A}_i, \overline{B}_i)$$

$$\overline{A}_i = W_i^{Ac} A$$

$$\overline{B}_i = W_i^{Bc} B$$

Appendix B: Geographical Proximity and Public Pension Fund Investments

Table B.1: Public Pension Investment Flows From Home State to Neighboring States

This table presents the results from OLS regressions using a single dummy variable. The sample consists of directed pairs of states in the U.S., including the District of Columbia, where 'directed' indicates a flow from a pension fund located in one state to another. The dependent variable is an indicator that takes a value of one if there was an investment from a public pension fund in one state to a VC fund situated in another state between 2001 and 2020, and zero otherwise. The independent variable, 'Proximate State N', is an indicator variable that assumes a value of one if a state that receives funding from a public pension is within the N^{th} closest states to the state where the public pension is located, where N represents the ordinal position, including the home state, and zero otherwise. t-statistics are presented in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

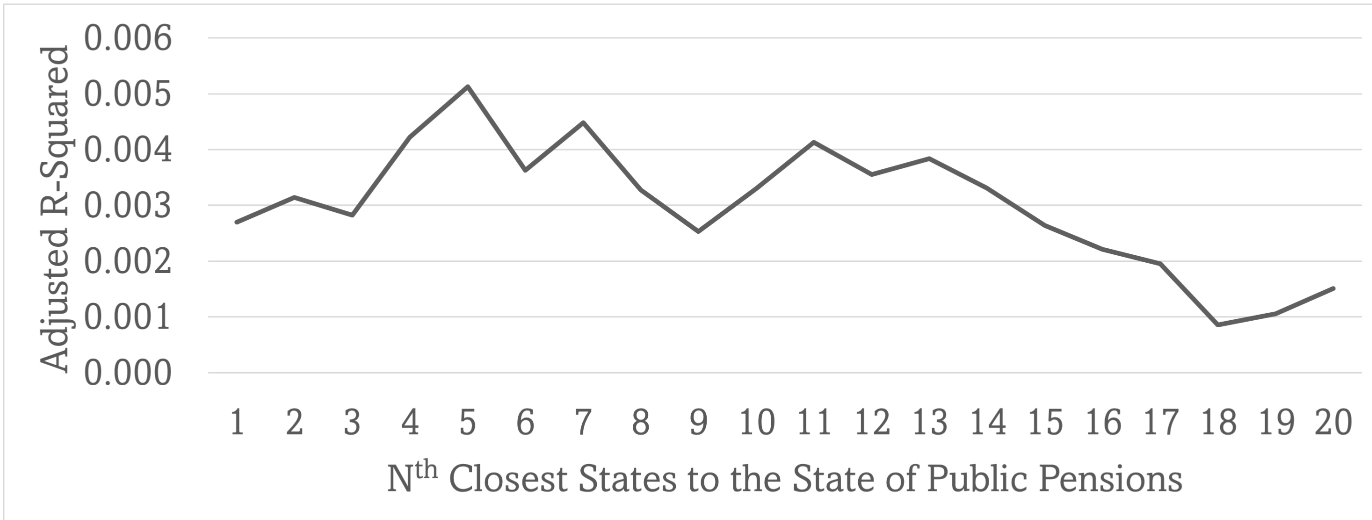
	Dep. Var. $I(Pension\ Investments)$				
Proximate State 1	0.075***				
	[2.93]				
Proximate State 2		0.071***			
		[3.14]			
Proximate State 3			0.062***		
			[2.99]		
Proximate State 4				0.068***	
				[3.59]	
Proximate State 5					0.070***
					[3.94]
Constant	0.107***	0.106***	0.106***	0.104***	0.102***
	[17.55]	[17.23]	[17.01]	[16.55]	[16.16]
R^2	0.305%	0.349%	0.318%	0.458%	0.549%
Adjusted R^2	0.270%	0.314%	0.282%	0.422%	0.513%
Obs	2809	2809	2809	2809	2809

	Dep. Var. $I(Pension\ Investments)$				
Proximate State 6	0.056***				
	[3.35]				
Proximate State 7		0.059***			
		[3.69]			
Proximate State 8			0.049***		
			[3.20]		
Proximate State 9				0.042***	
				[2.85]	
Proximate State 10					0.046***
					[3.21]
Constant	0.103***	0.102***	0.103***	0.103***	0.101***
	[16.11]	[15.70]	[15.66]	[15.55]	[15.14]
R^2	0.399%	0.483%	0.364%	0.289%	0.365%
Adjusted R^2	0.363%	0.448%	0.328%	0.253%	0.330%
Obs	2809	2809	2809	2809	2809

	Dep. Var. $I(Pension\ Investments)$				
Proximate State 11	0.050***				
	[3.55]				
Proximate State 12		0.045***			
		[3.32]			
Proximate State 13			0.046***		
			[3.44]		
Proximate State 14				0.042***	
				[3.21]	
Proximate State 15					0.037***
					[2.90]
Constant	0.100***	0.100***	0.099***	0.099***	0.100***
	[14.71]	[14.59]	[14.27]	[14.15]	[14.05]
R^2	0.448%	0.391%	0.420%	0.366%	0.299%
Adjusted R^2	0.413%	0.355%	0.384%	0.331%	0.264%
Obs	2809	2809	2809	2809	2809

	Dep. Var. $I(Pension\ Investments)$				
Proximate State 16	0.034***				
	[2.69]				
Proximate State 17		0.032**			
		[2.55]			
Proximate State 18			0.023*		
			[1.85]		
Proximate State 19				0.024**	
				[1.99]	
Proximate State 20					0.028**
					[2.29]
Constant	0.100***	0.101***	0.103***	0.102***	0.100***
	[13.93]	[13.75]	[13.93]	[13.60]	[13.18]
R^2	0.257%	0.231%	0.122%	0.141%	0.186%
Adjusted R^2	0.221%	0.195%	0.086%	0.106%	0.151%
Obs	2809	2809	2809	2809	2809

Figure B: Adjusted R^2 from Univariate Regression with Dummy Variables for States Ranked First to N-th in Proximity to a Public Pension Fund, Including Home State



Appendix C: Summary Statistics of the Instrumental Variable by State

The table presents the summary statistics of the instrumental variable used in the paper. The sample consists of investments made by US VC funds created between 2001 and 2017.

Fund State	Mean	SD	Fund State	Mean	SD	Fund State	Mean	SD
Alabama	0.325	0.470	Kentucky	0.000	0.000	North Dakota	0.000	0.000
Alaska	0.000	0.000	Louisiana	0.167	0.374	Ohio	0.206	0.405
Arizona	0.038	0.192	Maine	0.037	0.189	Oklahoma	0.147	0.355
Arkansas	0.294	0.462	Maryland	0.427	0.495	Oregon	0.135	0.342
California	0.257	0.437	Massachusetts	0.173	0.379	Pennsylvania	0.290	0.454
Colorado	0.170	0.375	Michigan	0.206	0.405	Rhode Island	0.000	0.000
Connecticut	0.238	0.426	Minnesota	0.398	0.490	South Carolina	0.099	0.300
D. of Columbia	0.292	0.455	Mississippi	0.000	0.000	South Dakota	0.333	0.478
Delaware	0.000	0.000	Missouri	0.329	0.470	Tennessee	0.180	0.384
Florida	0.166	0.372	Montana	0.000	0.000	Texas	0.140	0.347
Georgia	0.329	0.470	Nebraska	0.006	0.074	Unknown	0.000	0.000
Hawaii	0.000	0.000	Nevada	0.018	0.132	Utah	0.288	0.453
Idaho	0.330	0.473	New Hampshire	0.056	0.231	Vermont	0.155	0.363
Illinois	0.341	0.474	New Jersey	0.386	0.487	Virginia	0.000	0.000
Indiana	0.000	0.000	New Mexico	0.301	0.459	Washington	0.025	0.157
Iowa	0.024	0.152	New York	0.107	0.309	Wisconsin	0.238	0.426
Kansas	0.007	0.083	North Carolina	0.312	0.463	Wyoming	0.000	0.000

The Impact of Public Pension Funds as LPs on VC Investment

Jinyoung Kim

ONLINE APPENDIX

November 27, 2023

Part 1: Supplementary Analyses

Part 2: Technical Appendix

Online Appendix. Part 1: Supplementary Analyses

Public Pension Investors and startups' Exits: Linear Relationship

The table presents the linear relationships between the presence of public pension funds as LPs and the likelihood of exits. The sample consists of investments made by US VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. $I(Exit)$ is an indicator that assumes a value of one if a startup exits - either through acquisitions or IPOs - within fifteen years of the inception of a fund. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. $Avg(Log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and VC fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$I(Exit\_Acq)$			$I(Exit\_IPO)$			$I(Exit)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$I(PPFs)$	0.023**	0.035***	0.066***	0.052***	0.040***	0.036***	0.075***	0.075***	0.101***
	[2.02]	[4.33]	[7.30]	[8.14]	[7.17]	[7.06]	[7.66]	[11.71]	[13.19]
$Log(VC\ Firm\ Age)$			-0.065***			0.010***			-0.054***
			[-9.47]			[3.06]			[-7.50]
$Avg(Log(\#Deals))$			0			0			0.001
			[0.06]			[0.22]			[0.10]
Fund Focus×Founding Year FE		✓	✓		✓	✓		✓	✓
Fund Stage FE		✓	✓		✓	✓		✓	✓
$Adjusted\ R^2$	0	0.03	0.038	0.009	0.031	0.032	0.004	0.037	0.043
Observations	129154	129147	128934	129154	129147	128934	129154	129147	128934

Patent Startup Indicator and Exit Likelihood

The table presents the results of OLS regressions that explore the relationship between startup technologies and exit likelihood. The dependent variable $I(Exit)$ is an indicator that takes a value of one if a startup exits either through acquisitions or IPOs, zero otherwise. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. The sample includes startups with and without patent filings funded within the same period. The variable $PublicTechBase\_Avg$ represents the proportion of inventions by public firms in the patent references cited by startups before their exits. This is averaged across all of a venture's inventions during its tenure as a startup. $Log(RefIntensity)$ is the logarithm (plus one) of the number of patent references cited by the startup before an exit, normalized by the number of patent applications they submitted prior to that exit. $I(Patent)$ is an indicator variable that takes one if a venture has a patent before its exit, and zero otherwise. In the absence of patent data for some startups, I derive the values for $PublicTechBase\_Avg$ and $Log(RefIntensity)$ from similar startups possessing patent data, based on their business descriptions. These are denoted as $PublicTechBase\_Avg\_Imp$ and $Log(RefIntensity)\_Imp$, respectively. t-statistics, calculated using standard errors clustered at the intersection of founding year and economic sector are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1) $I(Exit)$	(2) $I(Exit\_Acq)$	(3) $I(Exit\_IPO)$	(4) $I(Exit)$	(5) $I(Exit\_Acq)$	(6) $I(Exit\_IPO)$
$I(Patent)$	0.087***	0.070***	0.017***	0.074***	0.070***	0.004
	[9.41]	[11.08]	[2.61]	[12.75]	[11.64]	[0.89]
$PublicTechBase\_Avg\_Imp$	0.216***	0.275***	-0.059**	0.211***	0.177***	0.034***
	[5.00]	[9.88]	[-2.34]	[10.93]	[9.67]	[2.86]
$Log(RefIntensity)\_Imp$	0.024***	0.014***	0.010***	0.031***	0.019***	0.012***
	[6.05]	[3.95]	[3.88]	[8.52]	[5.39]	[5.24]
Sector×Founding Year FE				✓	✓	✓
Adjusted R^2	0.012	0.014	0.001	0.069	0.023	0.104
Observations	42379	42379	42379	42178	42178	42178

Patenting Activities of Startups and Exit Likelihood

The table presents the results of OLS regressions that explore the relationship between startup patenting activity and the likelihood of exits. The dependent variable $I(Exit)$ is an indicator that takes a value of one if a startup exits either through acquisitions or IPOs, zero otherwise. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. The sample includes startups with and without patent filings funded within the same period. The variable $PublicTechBase\_Avg$ represents the proportion of inventions by public firms in the patent references cited by startups before their exits. This is averaged across all of a venture's inventions during its tenure as a startup. $Log(RefIntensity)$ is the logarithm (plus one) of the number of patent references cited by the startup before an exit, normalized by the number of patent applications they submitted prior to that exit. $Log(\#Applications + 1)$ indicates the count of distinct patent applications linked to a startup before its exit. In the absence of patent data for some startups, I derive the values for $PublicTechBase\_Avg$ and $Log(RefIntensity)$ from similar startups possessing patent data, based on their business descriptions. These are denoted as $PublicTechBase\_Avg\_Imp$ and $Log(RefIntensity)\_Imp$, respectively. t-statistics, calculated using standard errors clustered at the intersection of founding year and economic sector are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1) $I(Exit)$	(2) $I(Exit\_Acq)$	(3) $I(Exit\_IPO)$	(4) $I(Exit)$	(5) $I(Exit\_Acq)$	(6) $I(Exit\_IPO)$
$Log(\#Applications + 1)$	0.029***	0.013***	0.016***	0.031***	0.016***	0.016***
	[8.33]	[4.75]	[7.02]	[10.58]	[5.76]	[8.57]
$PublicTechBase\_Imp$	0.192***	0.251***	-0.059***	0.153***	0.160***	-0.007
	[6.90]	[11.76]	[-4.47]	[8.70]	[10.12]	[-0.88]
$Log(RefIntensity)\_Imp$	0.002	-0.001	0.003*	0.009***	0.004	0.005***
	[0.57]	[-0.34]	[1.91]	[3.03]	[1.38]	[3.34]
Sector×Founding Year FE				✓	✓	✓
Adjusted R^2	0.007	0.009	0.004	0.034	0.017	0.06
Observations	42379	42379	42379	42178	42178	42178

Effect of Public Pension Funds on Round of Funding

The table presents the impact of public pension funds on the number of funding rounds in which venture capital (VC) funds participate. The data set encompasses investments made by U.S.-based VC funds established between 2001 and 2017. The dependent variable is the number of funding rounds that a venture undergoes involving investment from a VC fund. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables representing each unique combination of the values of the variables specified the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables capturing every unique value combination from the specified columns. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
$I(\widehat{PPFs})$	2.255***	1.483***	1.041***
	[7.22]	[4.29]	[3.79]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.099	0.123	0.164
Observations	129067	128940	128823

Effect of Public Pension Funds on startups' Time to Exits

The table illustrates the influence of public pension funds on the duration of successful venture exits. The data set includes investments from U.S.-based venture capital funds founded between 2001 and 2017, which eventually led to successful venture exits. The dependent variable in this study is the time span, measured in years, from the establishment of the venture to its successful exit. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables representing each unique combination of the values of the variables specified the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables capturing every unique value combination from the specified columns. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
$I(\widehat{PPFs})$	2.144***	1.644**	0.81
	[3.61]	[2.30]	[1.47]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.152	0.167	0.212
Observations	37969	37871	37775

Effect of Public Pension Funds on the Number of Investors for an Investment

The table illustrates the influence of public pension funds on the number of funds participating in the same investment. The data set includes investments from U.S.-based venture capital funds founded between 2001 and 2017. The dependent variable is the number of funds participating in the same investment. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables representing each unique combination of the values of the variables specified the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables capturing every unique value combination from the specified columns. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
$I(\widehat{PPFs})$	-0.569***	-0.501**	-0.245
	[-2.71]	[-2.50]	[-1.50]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.067	0.091	0.12
Observations	129067	128940	128823

First-Stage Regressions with Firm Fixed Effects

The table presents the results from the conventional first-stage regressions in the presence of VC firm-fixed effects. The sample consists of investments made by US VC funds created between 2001 and 2017. The dependent variable is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise ($I(PPFs)$). In this regression, I include the instrumental variable, control variables, and fixed effects specified. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
Instrumental Variable	0.059***	0.058***	0.058***
	[3.13]	[2.95]	[3.00]
$log(VC\ Firm\ Age)$			-0.014*
			[-1.66]
$Avg(log(\#Deals))$			-0.025**
			[-2.34]
Firm FE	✓	✓	✓
Fund Focus$times$Founding Year FE	✓	✓	✓
Fund Stage FE		✓	✓
$Adjusted\ R^2$	0.748	0.75	0.75
Observations	151901	151901	151661

Effect of Public Pensions with Firm Fixed Effects

The table presents the effect of the presence of public pension funds as LPs on the main variables estimated from the traditional second-stage regressions. The sample consists of investments made by US VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. $I(\widehat{PPFs})$ represents the predicted value obtained from the first-stage regression. In this regression, the dependent variable is $I(PPFs)$, and it includes both the instrumental variable and all control variables, inclusive of fixed effects. In the second-stage regression, $I(\widehat{PPFs})$ and all other variables including fixed effects, with the exception of the instrumental variable, are treated as independent variables. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. $I(Exit)$ is an indicator that assumes a value of one if a startup exits - either through acquisitions or IPOs - within fifteen years of the inception of the VC fund. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
$I(\widehat{PPFs})$	0.181***	0.714***	-0.512***	0.628***
	[3.11]	[3.90]	[-3.02]	[3.54]
Controls	✓	✓	✓	✓
VC Firm FE	✓	✓	✓	✓
Fund Focus Vintage FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.15	0.118	0.211	0.159
Observations	100866	127924	127924	127924

	I(Exit_Acq)	I(Exit_IPO)	I(Exit)
$I(\widehat{PPFs})$	0.641***	0.111	0.752***
	[4.98]	[1.20]	[5.16]
Controls	✓	✓	✓
VC Firm FE	✓	✓	✓
Fund Focus×Founding Year FE	✓	✓	✓
Fund Stage FE	✓	✓	✓
Adjusted R^2	0.122	0.094	0.137
Observations	127924	127924	127924

	Time to Exit		
	All	Acquisitions	IPOs
$I(\widehat{PPFs})$	-10.021***	-9.490***	-9.946***
	[-6.50]	[-5.45]	[-3.32]
Controls	✓	✓	✓
VC Firm FE	✓	✓	✓
Fund Focus×Founding Year FE	✓	✓	✓
Fund Stage FE	✓	✓	✓
Adjusted R^2	0.598	0.55	0.542
Observations	39904	31339	5641

Public Pension Investors and VC Firm-Level Investment Decisions

The table indicates that public pension investors do not significantly alter the overall characteristics of VC firms. The sample is comprised of investments made by U.S. VC funds established between 2001 and 2017. The term $AfterPensionInv$ serves as an indicator variable: it takes the value of one if the period follows the VC firm's selection by a public pension plan and zero otherwise. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. $I(Exit)$ is an indicator that assumes a value of one if a startup exits - either through acquisitions or IPOs - within fifteen years of the investment. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
$AfterPensionInv$	-0.002	-0.047	-0.016	0.017
	[-0.20]	[-1.47]	[-0.59]	[0.68]
Controls	✓	✓	✓	✓
VC Firm FE	✓	✓	✓	✓
Adjusted R^2	0.128	0.088	0.139	0.087
Observations	35888	44630	44630	44630

	I(Exit_Acq)	I(Exit_IPO)	I(Exit)
$AfterPensionInv$	-0.02	0.023*	0.003
	[-0.63]	[1.78]	[0.08]
Constols	✓	✓	✓
VC Firm FE	✓	✓	✓
Adjusted R^2	0.098	0.07	0.105
Observations	44630	44630	44630

	Time to Exit		
	All	Acquisitions	IPOs
$AfterPensionInv$	0.029	0.057	0.201
	[0.06]	[0.11]	[0.41]
Constols	✓	✓	✓
VC Firm FE	✓	✓	✓
Adjusted R^2	0.3	0.315	0.316
Observations	15878	12460	3354

Influence of Public Pensions from the Same State on VC Investments

The table presents the decisions made by VC managers after their funds were selected by public pension investors located in the same state. The sample is comprised of investments made by U.S. VC funds established between 2001 and 2017. $I(SameStatePPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund from the same state as its LP and zero otherwise. $I(\widehat{SameState}PPFs)$ represents the predicted value obtained from the first-stage regression. In this regression, the dependent variable is $I(SameStatePPFs)$, and the model incorporates indicator variables that represent combinations of values for the instrumental variable, as well as all control variables. Notably, $I(PPFs)$ is not included in the regression. $I(\widehat{PPFs})$ is from a separate first-stage regression with indicator variables that represent combinations of values for the instrumental variable, as well as all control variables. The model does not include $I(SameStatePPFs)$. The second-stage regression include $I(\widehat{PPFs})$, $I(\widehat{SameState}PPFs)$ and indicator variables representing the intersection of all other variables, with the exception of the instrumental variable, as independent variables. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
$I(\widehat{SameState}PPFs)$	-0.002	-0.052	0.034	-0.078
	[-0.08]	[-0.66]	[0.63]	[-1.38]
$I(\widehat{PPFs})$	0.026**	0.189***	-0.121***	0.175***
	[2.01]	[2.83]	[-3.07]	[4.37]
Fund Focus	✓	✓	✓	✓
Founding Year	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓
Age Group	✓	✓	✓	✓
Size Group	✓	✓	✓	✓
Adjusted R^2	0.086	0.075	0.155	0.116
Observations	101673	128823	128823	128823

Effect of Public Pensions from the Same State on Exit Outcomes

The table presents the outcomes of the investments made by VC managers after their funds were selected by public pension investors located in the same state. The sample is comprised of investments made by U.S. VC funds established between 2001 and 2017. $I(SameStatePPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund from the same state as its LP and zero otherwise. $I(\widehat{SameStatePPFs})$ represents the predicted value obtained from the first-stage regression. In this regression, the dependent variable is $I(SameStatePPFs)$, and the model incorporates indicator variables that represent combinations of values for the instrumental variable, as well as all control variables. Notably, $I(PPFs)$ is not included in the regression. $I(\widehat{PPFs})$ is from a separate first-stage regression with indicator variables that represent combinations of values for the instrumental variable, as well as all control variables. The model does not include $I(SameStatePPFs)$. The second-stage regression include $I(\widehat{PPFs})$, $I(\widehat{SameStatePPFs})$ and indicator variables representing the intersection of all other variables, with the exception of the instrumental variable, as independent variables. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	I(Exit_Acq)	I(Exit_IPO)	I(Exit)
$I(\widehat{SameStatePPFs})$	-0.051	-0.023	-0.073
	[-1.04]	[-0.58]	[-1.18]
$I(\widehat{PPFs})$	0.066**	0.081***	0.147***
	[2.21]	[3.60]	[3.92]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group	✓	✓	✓
Size Group	✓	✓	✓
Adjusted R^2	0.075	0.072	0.087
Observations	128823	128823	128823
	Time to Exit		
	All	Acquisitions	IPOs
$I(\widehat{SameStatePPFs})$	-0.346	-0.174	-0.874
	[-0.50]	[-0.21]	[-0.94]
$I(\widehat{PPFs})$	-2.253***	-2.381***	-0.915**
	[-4.69]	[-4.74]	[-2.04]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group	✓	✓	✓
Size Group	✓	✓	✓
Adjusted R^2	0.414	0.437	0.433
Observations	40444	34187	6001

Effect of Public Pension Funds on Startup Exits - 10-Year Fund Life

The table shows the effect of public pension funds on the likelihood of exits, estimated from the second-stage regressions. The sample contains investments made by US VC funds created between 2001 and 2017. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables representing each unique combination of the values of the variables specified the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables capturing every unique value combination from the specified columns. $I(Exit)$ is an indicator that assumes a value of one if a startup exits - either through acquisitions or IPOs - within ten years of the fund's inception. $I(Exit\_Acq)$ is an indicator for exits via acquisition, and $I(Exit\_IPO)$ is an indicator for IPO exits. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$I(Exit\_Acq)$			$I(Exit\_IPO)$			$I(Exit)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$I(\widehat{PPFs})$	0.140***	0.138***	0.092***	0.105***	0.082***	0.070***	0.245***	0.220***	0.162***
	[5.42]	[4.89]	[4.40]	[5.26]	[5.14]	[4.69]	[6.87]	[6.41]	[6.90]
Fund Focus	✓	✓	✓	✓	✓	✓	✓	✓	✓
Founding Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage	✓	✓	✓	✓	✓	✓	✓	✓	✓
Age Group		✓	✓		✓	✓		✓	✓
Size Group			✓			✓			✓
Adjusted R^2	0.039	0.065	0.079	0.04	0.059	0.075	0.041	0.068	0.085
Observations	129067	128940	128823	129067	128940	128823	129067	128940	128823

Effect of Public Pension Funds on the Time to Exit in VC Investments - 10-Year Fund Life

The table presents the impact of public pension funds on investment exit timing, as estimated from the second-stage regression analyses. The sample comprises investments made by U.S.-based VC funds established between 2001 and 2017 that ultimately achieved successful exits. The dependent variable is the duration, measured in years, between the founding year of the VC fund and the occurrence of a successful investment exit, limited to exits that take place within ten years of the fund's inception. $I(PPFs)$ is an indicator variable that equals one if a VC fund has a public pension fund as its LP and zero otherwise. In the first-stage regressions, $I(PPFs)$ is regressed on the interaction terms between the instrumental variable and a set of indicator variables representing each unique combination of the values of the variables specified the columns. In the second-stage regressions, the dependent variable is regressed on the predicted value of $I(PPFs)$ ($I(\widehat{PPFs})$) from the first-stage regression and a set of indicator variables capturing every unique value combination from the specified columns. Age group categorizes VC firms into three equal groups based on the firm's age, with divisions made annually. Size group divides VC firms into three equal segments based on the average number of deals their previous funds invested in, also categorized annually. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

All exits	(1)	(2)	(3)
$I(\widehat{PPFs})$	-0.818***	-0.563***	-0.578***
	[-3.71]	[-3.00]	[-2.90]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.224	0.246	0.279
Observations	28155	28067	27980

Acquisitions	(1)	(2)	(3)
$I(\widehat{PPFs})$	-0.825***	-0.590**	-0.674***
	[-4.27]	[-2.55]	[-3.00]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.234	0.256	0.291
Observations	23499	23420	23329

IPOs	(1)	(2)	(3)
$I(\widehat{PPFs})$	-0.296	-0.093	0.195
	[-0.60]	[-0.22]	[0.58]
Fund Focus	✓	✓	✓
Founding Year	✓	✓	✓
Fund Stage	✓	✓	✓
Age Group		✓	✓
Size Group			✓
Adjusted R^2	0.276	0.308	0.367
Observations	4595	4542	4485

Exit Strategies and Round-to-Round Returns

The table presents the relationship between round-to-round returns on investments that are known to exit through particular exit methods. The sample includes investments by U.S.-based VC funds from 1980 to 2022 with available round-to-round returns. The dependent variable represents returns between the current funding round and the subsequent one. $I(Exit)$ is an indicator variable set to one when the investment exits during the fund's lifespan. $I(Exit_IPO)$ is set to one if the investment undergoes an IPO exit within this period. Similarly, $I(Exit_Acq)$ is set to one for exits via acquisition within the fund's duration. $log(VC\ Firm\ Age)$ is the log of one plus the age of the VC firm. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)
$I(Exit\_Acq)$	0.008			0.179***
	[0.22]			[4.31]
$I(Exit\_IPO)$		0.437***		0.506***
		[7.84]		[8.36]
$I(Exit)$			0.324***	
			[7.49]	
$log(VC\ Firm\ Age)$	-0.102**	-0.094*	-0.093*	-0.091*
	[-2.03]	[-1.88]	[-1.87]	[-1.83]
$Avg(log(\#Deals))$	0.072***	0.070***	0.070***	0.069***
	[3.41]	[3.28]	[3.30]	[3.26]
Fund Focus×Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.005	0.008	0.007	0.008
Observations	31466	31466	31466	31466

Presence of Public Pensions and Round-to-Round Returns

The table presents the relationship between the presence of public pension investors and round-to-round returns. The sample includes investments by U.S.-based VC funds created since 2001 with available round-to-round returns. The dependent variable represents returns between the current funding round and the subsequent one. $I(PPFs)$ is an indicator variable that takes a value of one if a VC fund has public pensions as its investors, zero otherwise. $log(VC\ Firm\ Age)$ is the log of one plus the age of the VC firm. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)
$I(PPFs)$	-0.481***	0	-0.027
	[-3.86]	[-0.01]	[-0.43]
$log(VC\ Firm\ Age)$			-0.026
			[-0.31]
$Avg(log(\#Deals))$			0.138***
			[2.88]
Fund Focus×Founding Year FE		✓	✓
Fund Stage FE		✓	✓
Adjusted R^2	0.001	0.032	0.037
Observations	6700	3179	3105

Relationship between Pension Investors' Funded Ratio and VC Investments

The table shows that the linear relationship between the weighted average of funded ratio of public pensions and VC managers' investment decisions. The sample is comprised of investments made by U.S. VC funds with public pension investors established between 2001 and 2017. The weighted average funded ratio is computed using the funded ratio of public pension investors and the weights determined by the total assets of public pensions. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
Weighed Average of Funded Ratio	-0.031	-0.066	0.03	-0.01
	[-1.51]	[-0.91]	[1.15]	[-0.31]
Fund Focus×Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.056	0.04	0.117	0.052
Observations	19687	24809	24809	24809

Relationship between Pension Investors' Returns and VC Investments

The table shows that the gap between assumed and actual returns of public pensions does not significantly affect VC managers' investment decisions. The sample is comprised of investments made by U.S. VC funds with public pension investors established between 2001 and 2017. I compute the weighted average difference between the assumed rate of returns and the investment returns from the past three years. The weights are based on the total assets of the public pensions. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
Gap in Assumed and Actual Returns	0.090***	0.042	0.017	-0.07
	[2.73]	[0.25]	[0.18]	[-0.88]
Fund Focus×Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.055	0.041	0.123	0.054
Observations	18721	23592	23592	23592

Relationship between Funded Ratio and VC Investments - Pension-Investment Pairs

The table shows that the linear relationship between the funded ratio of public pensions and VC investments. The sample is comprised of pairs of pension investment units and VC investments made by U.S. VC funds with public pension investors established between 2001 and 2017. Funded ratios are collected from Public Plans Data database. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus,fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	PublicTechBase		log(#Patents)		Early Stage		Recent Late Round Funding	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Funded Raio	-0.004	-0.003	-0.001	0.023	0.002	0.014	-0.002	-0.018
	[-1.60]	[-0.53]	[-0.07]	[1.12]	[0.40]	[1.05]	[-0.46]	[-1.17]
Log(VC Firm Age)	-0.031***	-0.030***	0.749***	0.753***	-0.359***	-0.361***	0.555***	0.558***
	[-2.83]	[-2.79]	[8.89]	[8.91]	[-8.86]	[-8.89]	[10.12]	[10.17]
Avg(Log(#Deals))	0.026***	0.025***	0	-0.001	-0.023	-0.022	0.049	0.047
	[3.20]	[3.16]	[0.01]	[-0.02]	[-0.79]	[-0.76]	[1.35]	[1.34]
Fund Focus×Founding Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
VC Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Pension State FE		✓		✓		✓		✓
Adjusted R^2	0.136	0.136	0.115	0.116	0.174	0.175	0.117	0.117
Observations	67315	67314	94218	94218	94218	94218	94218	94218

Relationship between Assumed and Actual Return Gap and VC Investments - Pension-Investment Pairs

The table shows that the linear relationship between the gap between assumed and actual returns of public pensions and VC investments. The sample is comprised of pairs of pension investment units and VC investments made by U.S. VC funds with public pension investors established between 2001 and 2017. I compute the difference between the reported assumed rate of returns and the investment returns from the past three years. *PublicTechBase_Imp* denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, fund stage, and founding year are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	PublicTechBase		log(#Patents)		Early Stage		Recent Late Round Funding	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Return Gap	0.001	0.003	0.023*	0.032**	-0.026**	-0.021*	0.011	-0.002
	[0.17]	[0.56]	[1.67]	[1.98]	[-2.44]	[-1.66]	[0.80]	[-0.13]
Log(VC Firm Age)	-0.059***	-0.058***	0.838***	0.840***	-0.364***	-0.367***	0.608***	0.612***
	[-5.02]	[-4.90]	[6.98]	[6.97]	[-7.38]	[-7.44]	[8.98]	[9.09]
Avg(Log(#Deals))	0.033**	0.032**	0.101	0.097	-0.109**	-0.105**	0.125	0.118
	[2.21]	[2.15]	[1.43]	[1.44]	[-2.20]	[-2.22]	[1.61]	[1.62]
Fund Focus×Founding Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
VC Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Pension State FE		✓		✓		✓		✓
Adjusted R^2	0.142	0.143	0.139	0.139	0.222	0.222	0.131	0.131
Observations	42981	42979	62274	62274	62274	62274	62274	62274

Heterogeneity in Public Pension Funds' Decisions - Different Criteria

The table shows which characteristics of previous funds are considered by public pension funds based on their funded status and past investment returns. The dataset includes public pension-backed VC funds established between 2001 and 2021. These funds had a history of investments from previous VC funds spanning three to six years before their creation. In the first four columns, the dependent variable is an indicator variable that takes a value of one if the fund has a public pension as an investor whose funded ratio is below one, and zero otherwise. In the subsequent columns, the dependent variable is set to one if the fund has a public pension investor with a difference between its Assumed Ratio of Returns (ARR) and the investment return of the past three years exceeding 0. Otherwise, it's set to zero. $PublicTechBase\_Past$ is the average value of $PublicTechBase\_Imp$ for startups invested by previous funds, within a timeframe of three to six years before the establishment of the focal fund. Similarly, Revised: $log(\#Patents)\_Past$ represents the mean of the natural logarithm of the total number of patents (with an addition of one) from past investments. $R(Early\_Inv)_Past$ denotes the share of early-stage investments conducted by these previous funds during the same three- to six-year timeframe. Likewise, $R(Recent\_Late\_Round)\_Past$ indicates the proportion of late-round investments with a history of another late round within the past two years executed by these previous funds. $Avg(Log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of fund focus, fund stage, and founding year levels, are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	Invested by Underfunded Pensions				Invested by Pensions Not Keeping Up with ARR			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$PublicTechBase\_Past$	0.043				0.097			
	[0.41]				[0.83]			
$log(\#Patents)\_Past$		0				-0.005		
		[0.02]				[-0.11]		
$R(Early\_Inv)_Past$			-0.062				-0.151*	
			[-0.97]				[-1.97]	
$R(Recent\_Late\_Round)\_Past$				0.131				0.131
				[1.25]				[1.36]
$Avg(Log(\#Deals))$	0.035***	0.036***	0.039***	0.036***	0.012	0.02	0.019	-0.001
	[3.64]	[3.46]	[3.92]	[3.67]	[0.46]	[0.71]	[0.80]	[-0.02]
$Log(VC\ Firm\ Age)$	-0.022	-0.019	-0.019	-0.039**	0.03	0.032*	0.038*	0.031
	[-1.39]	[-1.04]	[-1.31]	[-2.22]	[1.61]	[1.73]	[1.98]	[1.66]
Fund Focus×Founding Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.238	0.238	0.239	0.242	0.559	0.558	0.561	0.56
Observations	1133	1133	1133	1133	1133	1133	1133	1133

Reduced-Form Regressions - Investment Choices

The table displays the results of reduced form regressions utilizing two instrumental variables: one as defined in the paper and another created by altering the definition of the vintage year. The latter redefines the vintage year as the first year of a VC fund's investments, diverging from the original use of the founding year to construct the instrumental variable. The sample includes investments by U.S.-based VC funds from 1980 to 2022 with a founding year between 2001 and 2017. $PublicTechBase\_Imp$ denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. Other dependent variables include the logarithm of the granted patents count as of the investment date (increased by one), a dummy variable indicating early-stage deals, and an indicator for startups that have secured late-round capital within the preceding two years. $log(VC\ Firm\ Age)$ is the log of one plus the age of the VC firm. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Panel A: Instrumental variable as in the paper

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
Instrumental Variable	0.023***	0.076***	-0.063***	0.082***
	[4.62]	[5.31]	[-5.42]	[7.19]
$log(VC\ Firm\ Age)$	-0.005*	0.065***	-0.090***	0.081***
	[-1.85]	[7.29]	[-12.21]	[9.67]
$Avg(log(\#Deals))$	0.002	-0.028***	0.047***	-0.024***
	[0.95]	[-6.97]	[7.11]	[-4.38]
Fund Focus×Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.05	0.044	0.119	0.08
Observations	101827	128934	128934	128934

Panel B: Instrumental variable created using the first year of investments as vintage

	$PublicTechBase\_Imp$	log(#Patents)	Early Stage	Recent Late Round
Instrumental Variable	0.023***	0.043***	-0.059***	0.066***
	[4.84]	[3.20]	[-4.29]	[4.96]
$log(VC\ Firm\ Age)$	-0.005**	0.071***	-0.088***	0.082***
	[-2.09]	[7.64]	[-11.73]	[9.18]
$Avg(log(\#Deals))$	0.002	-0.029***	0.047***	-0.024***
	[0.93]	[-6.74]	[7.07]	[-4.27]
Fund Focus×Founding Year FE	✓	✓	✓	✓
Fund Stage FE	✓	✓	✓	✓
Adjusted R^2	0.05	0.043	0.119	0.079
Observations	101827	128934	128934	128934

Reduced-Form Regressions - Exit Outcomes

The table displays the results of reduced form regressions utilizing two instrumental variables: one as defined in the paper and another created by altering the definition of the vintage year. The latter redefines the vintage year as the first year of a VC fund's investments, diverging from the original use of the founding year to construct the instrumental variable. The sample includes investments by U.S.-based VC funds from 1980 to 2022 with a founding year between 2001 and 2017. $I(Exit\_IPO)$ is set to one if the investment undergoes an IPO exit within this period. Similarly, $I(Exit\_Acq)$ is set to one for exits via acquisition within the fund's duration. The dependent variable in the last column is the time to exit measured in years from the inception of a VC fund. $log(VC\ Firm\ Age)$ is the log of one plus the age of the VC firm. $Avg(log(\#Deals))$ is the average number of investments made by previous funds. t-statistics, calculated using standard errors clustered at the intersection of VC fund focus, founding year, and fund stage are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Panel A: Instrumental variable as in the paper			
	$I(Exit\_Acq)$	$I(Exit\_IPO)$	Time to Exit
Instrumental Variable	0.074***	0.041***	-1.943***
	[7.58]	[6.80]	[-11.34]
$log(VC\ Firm\ Age)$	-0.067***	0.007**	0.798***
	[-10.90]	[2.13]	[7.62]
$Avg(log(\#Deals))$	0.007	0.003*	0.339***
	[1.59]	[1.79]	[3.45]
Fund Focus×Founding Year FE	✓	✓	✓
Fund Stage FE	✓	✓	✓
Adjusted R^2	0.039	0.034	0.343
Observations	128934	128934	40617
Panel B: Instrumental variable created using the first year of investments as vintage			
	$I(Exit\_Acq)$	$I(Exit\_IPO)$	Time to Exit
Instrumental Variable	0.085***	0.042***	-2.144***
	[8.69]	[7.30]	[-13.76]
$log(VC\ Firm\ Age)$	-0.073***	0.005	0.935***
	[-11.11]	[1.62]	[8.91]
$Avg(log(\#Deals))$	0.007*	0.003*	0.331***
	[1.68]	[1.80]	[3.33]
Fund Focus×Founding Year FE	✓	✓	✓
Fund Stage FE	✓	✓	✓
Adjusted R^2	0.039	0.034	0.346
Observations	128934	128934	40617

Selection and Influence by VC Managers on Technological Focus

In this section, I unpack the positive effect of public pension investors on the startups' technological alignment with public companies, focusing on two key avenues: venture selection by managers and their direct influence on portfolio startups. Each of these mechanisms is related to the technological orientation of portfolio startups.

The first mechanism centers on the venture selection process undertaken by managers. Previous research has highlighted that VCs rely heavily on rigorous selection procedures when curating their portfolio, often considering deal selection as the pivotal factor for determining the success or failure of an investment (Gompers et al. (2020)). This process usually involves comprehensive information collection from entrepreneurs and the application of advanced predictive algorithms, such as machine learning, to anticipate the future performance of startups (Bonelli (2023)). Should managers perceive that public pension funds favor stable returns over high-risk, high-reward opportunities, they may be inclined to assemble a portfolio skewed towards startups with more predictable outcomes.

The second mechanism posits that VC managers exert a more direct influence over their portfolio companies (Hellmann and Puri (2002), Fitza, Matusik, and Mosakowski (2009), Bernstein, Giroud, and Townsend (2016)). In this scenario, managers actively guide their startups to explore technologies that align with the interests of public firms. This guidance occurs even if the startups in the original portfolio do not inherently focus on technologies comparable to those of public companies. By doing so, VCs enhance the startups' attractiveness to public firms, thereby increasing the likelihood of acquisition deals.

To investigate the mechanisms that underlie the impact of VCs with pension investors on the technological focus of startups, I assemble a sample comprised of venture-year pairs spanning the venture's life as a startup. The dependent variable, $PublicTechBase\_Imp$, serves to quantify the technological attributes of these startups.

Two key dummy variables are examined in this analysis. The first variable is concerned with startups that have been selected for investment by VC funds with public pension investors. This variable is assigned a value of one if the venture is financed by a VC fund with public pension backing, and zero otherwise. This variable aims to capture the average characteristics of startups that attract investment from VC funds supported by public pensions.

The second variable focuses on the periods following the investment by VC funds with public pension involvement. It assumes a value of one for time periods subsequent to such an investment, and zero otherwise. If a venture secured investments from VC funds with pension investors on multiple occasions, I consider the earliest instance to create the variable. This variable is designed to capture any potential shifts in the technological focus of portfolio startups that may be influenced by VCs with public pension fund backing.

The table below reports the results from OLS regressions with various fixed effects. Across all model specifications, both indicator variables display positive and statistically significant coefficients. This suggests that the technological attributes of startups targeted by VCs with public pension investors are likely to be aligned with those of public firms even before being invested by those VCs. Additionally, VCs appear to guide their portfolio companies in a direction that enhances the likelihood of acquisitions. In column (4), where fixed effects for the venture's economic sector and year combinations are included, the coefficient for the indicator variable representing the post-investment period by VC funds with public pension investors is larger in magnitude than that of the indicator for startups selected by those VC funds. This indicates the active role that these VC funds play in influencing their portfolio startups' technological focus.

The choice to invest in startups that focus on technologies aligned with public firms suggests that VC funds backed by public pensions are likely to support startups with a good chance of exit through acquisitions. Furthermore, the analysis indicates that managers can also steer these startups to adopt technologies that more closely align with those of public firms.

Selection and Influence by VC Managers on Technological Focus

The table shows that managers of VC funds with public pensions both select startups focused on technology similar to that of public firms and steer them to pursue that direction. The sample contains venture-year pairs throughout the venture's tenure as a startup. The dependent variable is $PublicTechBase\_Imp$, which denotes the fraction of patents from public firms cited by a venture in the latest year of its patent publication. For startups without patent data, I derive the value from analogous startups, matching based on business descriptions and investment years. There are two dummy variables. The first pertains to startups selected by VC funds with public pension funds. The second relates to periods post-investment by VC funds with public pension funds. t-statistics, calculated using standard errors clustered at the intersection of economics sector and year levels, are presented in brackets. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

	(1)	(2)	(3)	(4)
startups Selected by VC Fund with Pensions	0.035***	0.028***	0.022***	0.016***
	[5.32]	[6.35]	[2.77]	[3.29]
Post Investment by VC Fund with Pensions	0.031***	0.028***	0.042***	0.039***
	[4.35]	[5.66]	[4.60]	[7.77]
Venture Sector FE		✓		
Year FE			✓	
Venture Sector×Year FE				✓
Adjusted R^2	0.02	0.176	0.032	0.19
Observations	230777	230750	230777	230766

Share of Early-Stage Investments by Fund Stage

The table presents the proportion of early-stage investments (referring to both seed and early stage investments in the paper), categorized by fund stage, using classifications from VentureXpert. The sample includes U.S. VC-backed investments from 1980 to 2022.

Fund_Stage	Obs	Share
Seed Stage	21,322	0.596
Early Stage	143,216	0.430
Balanced Stage	205,103	0.388
Energy	1,686	0.101
Fund of Funds	3,495	0.190
Generalist	42,831	0.187
Later Stage	28,775	0.190

Online Appendix. Part 2: Technical Appendix

I provide an extensive overview of the data process applied in this paper. In Section 1, I detail the method used for merging two databases using company names. Section 2 covers the methodology used to identify patent publications associated with public companies. In Section 3, I provide a list comprising all the airports I employ in my research to calculate the distance between two states.

1. Tiered, Exact Name Matching

There are cases where I combine two distinct databases by name of companies. I explain in this section, the approach I take with an example where I find patent publications of a startup.

For each name of startups in VentureXpert and the first assignees in Google Patents, I create three distinct name variables. The first variable, 'Name 1', is derived by replacing "&amp" with "&" in the original name, ensuring minimal alteration. The second variable, 'Name 2', is generated by removing punctuation marks, parentheses, vertical lines, backslashes, and forward slashes from 'Name 1'. Lastly, 'Name 3' is created by eliminating regional information and legal suffixes from 'Name 2'.

The rationale behind these three variables is to prioritize matching by 'Name 1', which preserves the original name's integrity. If unmatched, I proceed with 'Name 2', which has undergone further modification to enhance the matching potential. Finally, for any remaining unmatched names, I utilize 'Name 3'. This tiered approach aims to maintain matching quality while maximizing the number of matched names.

2. Patent Publications of Public Firms

In order to identify patent publications linked to public companies, I generate two datasets based on the names of these companies.

1. CRSP Name History

   CRSP stocknames file contains PERMCO, Company names, the beginning dates and end dates of name effective date, and other variables. I keep PERMCO, Company names, and the beginning dates and end dates of name effective date, and other variables.

2. CIK Name History

> This dataset is derived from a combination of SEC bulk data on company name history and Compustat name header files. The Compustat files are utilized to identify the CIKs of companies required to submit 10K reports, as well as the first and last years of a firm's reporting history within the file. The SEC bulk data on historical company names, however, does not include the first day when a CIK became available or the last day it appeared in the system. To address this limitation, the Compustat files are employed to infer the first and last effective dates of a company's name. Additionally, I supplemented missing name effective dates with historical information whenever necessary. The final dataset comprises GVKEY, CIK, company names from the SEC file, and the effective start and end dates of those names.

I begin the process with public companies from the CRSP dataset by applying the tiered, exact name matching explained above. In the initial matching process, I used the CRSP Name History dataset and the 'Name1' variable, along with date information (using the 'publication date' from the patent dataset and the start and end dates of name effective date). I identified 2,200,725 patent publications. Moving on to the 'Name2' variable, I found 637 additional patent publications. Lastly, when matching by the 'Name3' variable, I obtained 206,167 more matches. During this process, I tallied the number of 'permco' for each publication-publication date pairing. I retained the publication-publication date-permco entries with only one matching 'permco', ensuring no publication-publication date instances had two 'permco' matches.

Next, for the remaining unmatched patent publications, I utilized another dataset for public company names, 'CIK Name History'. Initially, I matched by 'Name1', incorporating date information from Compustat and 'publication date', which resulted in 784,384 additional matches. With 'Name2', I discovered 46,156 more matches. Finally, using 'Name3', I obtained 87,508 further matched patent publications. In each step, I also ensure that there is a unique CIK for every matched publication.

The process results in 3,325,577 unique patent publications for public companies available in the CRSP and Compustat databases.

3. Major Airports by State

I use data from the Bureau of Transportation Statistics on inter-airport distances to estimate likely travel distances between states for General Partners (GPs) and Limited Partners (LPs). This process involves designating a specific major airport for each state, with San Francisco Airport chosen for California due to its proximity to

Silicon Valley. The list includes all the airports that I utilize to calculate the distance between two states in my research.

- Alabama: Birmingham-Shuttlesworth International (BHM)
- Alaska: Ted Stevens Anchorage International (ANC)
- Arizona: Phoenix Sky Harbor International (PHX)
- Arkansas: Bill and Hillary Clinton Nat Adams Field (LIT)
- California: San Francisco International (SFO)
- Colorado: Denver International (DEN)
- Connecticut: Bradley International (BDL)
- Delaware: New Castle (ILG)
- Florida: Orlando International (MCO)
- Georgia: Hartsfield-Jackson Atlanta International (ATL)
- Hawaii: Daniel K Inouye International (HNL)
- Idaho: Boise Air Terminal (BOI)
- Illinois: Chicago O'Hare International (ORD)
- Iowa: Des Moines International (DSM)
- Kansas: Wichita Dwight D Eisenhower National (ICT)
- Kentucky: Cincinnati/Northern Kentucky International (CVG)
- Louisiana: Louis Armstrong New Orleans International (MSY)
- Maine: Portland International Jetport (PWM)
- Maryland: Baltimore/Washington International Thurgood Marshall (BWI)
- Massachusetts: Logan International (BOS)
- Michigan: Detroit Metro Wayne County (DTW)
- Minnesota: Minneapolis-St Paul International (MSP)
- Mississippi: Jackson Medgar Wiley Evers International (JAN)
- Missouri: St Louis Lambert International (STL)
- Montana: Bozeman Yellowstone International (BZN)
- Nebraska: Eppley Airfield (OMA)
- Nevada: Harry Reid International (LAS)
- New Hampshire: Manchester Boston Regional (MHT)
- New Jersey: Newark Liberty International (EWR)
- New Mexico: Albuquerque International Sunport (ABQ)
- New York: John F. Kennedy International (JFK)
- North Carolina: Charlotte Douglas International (CLT)
- North Dakota: Minot International (MOT)
- Ohio: Cleveland-Hopkins International (CLE)

- Oklahoma: Will Rogers World (OKC)
- Oregon: Portland International (PDX)
- Pennsylvania: Philadelphia International (PHL)
- Rhode Island: Rhode Island Tf Green International (PVD)
- South Carolina: Charleston AFB/International (CHS)
- South Dakota: Joe Foss Field (FSD)
- Tennessee: Nashville International (BNA)
- Texas: Dallas/Fort Worth International (DFW)
- Utah: Salt Lake City International (SLC)
- Vermont: Burlington International (BTV)
- Virginia: Washington Dulles International (IAD)
- Washington: Seattle/Tacoma International (SEA)
- West Virginia: West Virginia International Yeager (CRW)
- Wisconsin: General Mitchell International (MKE)
- Wyoming: Casper/Natrona County International (CPR)

2 New Technologies and Stock Returns

New technology is an engine of the growth of an economy. However, developing it is not without risks. For example, many companies, including IBM, Google, and Microsoft, are increasingly investing in quantum computing technology that could drastically change our lives. However, experts say it is uncertain how long it will take to build a fully-fledged quantum computer. Some even question the practical viability of the technology itself. This example underscores that investing in new technologies and, in turn, investing in stocks of firms that heavily invest in new technologies is highly risky.

This paper poses a simple question: Do investments in stocks of firms with high exposure to new technologies lead to potentially higher returns because of their association with uncertainty that increases idiosyncratic risks?

The normative reasoning underpinning the risk-return trade-off has been extensively addressed by modern portfolio theories (Sharpe (1964), Lintner (1975), Black (1972), Merton (1973), to name a few). The risks are often defined as the covariance between the returns of individual securities and market-wide factors. Idiosyncratic risks are not priced since investors face minimal constraints in buying or shorting multiple stocks, allowing them to hold a diversified portfolio and mitigate the impact of firm-specific risks. Although the theories provide a holistic view where investors consider all market securities and their wealth when investing, in reality, many end up with portfolios heavily tilted towards certain stocks (Goetzmann, Kumar et al. (2005)). In that case, the primitive concept of risk-return trade-off suggested by Von Neumann and Morgenstern (2007) might be reflected in security prices. In other words, under-diversified investors may still require a higher risk premium for holding stocks of companies that have more uncertain cash flows (Levy (1978), Merton et al. (1987)).

Prior empirical results, however, find a negative relationship between historical idiosyncratic volatility and subsequent stock returns (Ang et al. (2006)). Fu (2009) emphasizes the importance of employing a metric that captures expected idiosyncratic risks over realized historical volatility to observe a positive relation. To the extent that idiosyncratic risks exhibit some consistent patterns within a firm, constructing a proxy for expected idiosyncratic risks from historical return volatility helps examine the idiosyncratic risk-return relationship.

Instead of relying on statistical assumptions on the properties of idiosyncratic risks, I utilize a firm characteristic that relates to the uncertainty of future cash flows.

Developing emerging technologies often brings inherent uncertainty and ambiguity to a firm (Rotolo, Hicks, and Martin (2015)). For instance, accurately predicting the amount of corporate resources required for developing new technology can be challenging. Furthermore, success is not always guaranteed. Yet, if the technology succeeds commercially, it can create significant market demand and rewards for the company.

Leveraging this insight, I investigate the potential effect of developing new technologies on future stock returns, anticipating a positive relationship in markets where investors require a risk premium for idiosyncratic risks.

I begin by identifying firms with high exposure to new technologies, specifically those that are actively involved in innovative activities related to emerging technologies. I compile data on U.S. patent publications from Google Patents, covering the years 1976 to 2021. Additionally, I gather neighboring patents from various countries, which either cite or are cited by these U.S. patent publications. Noting that a single patent publication can be publicized multiple times by the patent office, I focus on the first instance of public availability for each patent publication. This initial instance is referred to as the 'invention.' By utilizing a set of unique inventions, I develop a deep learning model that incorporates both text and network data for each invention. This model generates invention-level embeddings, which are then used to create technology clusters at the end of June each year. New tech clusters are identified based on the information pertaining to the inventions within these clusters. Specifically, I define new tech clusters as those experiencing significant growth in young inventions. Subsequently, I use a firm's past three-year inventions to calculate its exposure to emerging technologies, measured as the fraction of inventions belonging to new tech clusters.

To ensure that inventions classified as new technologies correspond with our existing knowledge of such technologies, I investigate the attributes of these inventions. I find that an invention is more likely to be part of a new technology cluster if it possesses certain characteristics: it cites younger patents, references a higher number of non-patent sources, originates from public firms and their subsidiaries, and frequently cites patents from these companies. This pattern suggests that inventions within new technology clusters tend to leverage up-to-date and diverse knowledge sources, reflecting their cutting-edge nature. Furthermore, the technological reliance on public firms and their subsidiaries aligns with their crucial roles in innovation, primarily due to their substantial R&D investments and broad collaborative networks.

I further examine the firm-level exposure to new technology in relation to existing innovation variables introduced by previous studies: a measure of innovative originality of a firm introduced by Hirshleifer, Hsu, and Li (2018), a measure of a firm's ability to efficiently generate innovative work introduced by Hirshleifer, Hsu, and Li (2013), a measure of R&D intensity (Chan, Lakonishok, and Sougiannis (2001)), a significantly positive abnormal increase in R&D expenditures defined by Eberhart, Maxwell, and Siddique (2004), and a measure of a firm's ability to turn R&D expenditures into future sales introduced by Cohen, Diether, and Malloy (2013). I find that firms' originality and a huge growth in R&D have a positive relationship with the exposure. However, firms' ability to efficiently produce innovation or firms' ability to turn R&D into sales is negatively associated with the exposure to new technologies, showing that exploring emerging technologies is far from prioritizing efficiency.

Leveraging the firm-level exposure to emerging technologies, I conduct a portfolio analysis. Specifically, I construct value-weighted portfolios based on firms' exposure to new technologies at the end of June for each year, spanning from 1981 to 2019. Four distinct portfolios are created: stocks with zero exposure to new technologies and three groups of stocks with positive exposure to new technologies ('Low', 'Middle', 'High'). The three positive-exposure portfolios are formed according to the 30th and 70th percentiles of the exposure distribution. The four portfolios are maintained and evaluated over a subsequent twelve-month holding period (July of year t to June of year $t+1$). The 'High' portfolio outperforms other portfolios in terms of excess returns, both in single-sort and double-sort analyses, with the latter employing the NYSE median size breakpoint for further stock classification. The zero-cost 'High'-'Low' portfolio yields an average monthly return of 0.0058 (t=3.43) throughout the sample period. A size-adjusted portfolio is constructed from double-sorted (size and new tech exposure) portfolios. This portfolio generates an average monthly return of 0.006 (t=2.99), equivalent to an annual return of 7.4%. This portfolio is referred to as NMO (new minus old).

A comprehensive examination of NMO's alphas across various factor models is conducted. The alpha obtained from the Carhart (1997) model amounts to 0.0063 (t=3.41), corresponding to a 7.83% annualized alpha. The alpha derived from the model incorporating four factors, plus the robust-minus-weak (RMW) and conservative-minus-aggressive (CMA) factors, is 0.0096 (t=5.29), equating to a 12.15% annualized alpha. Additionally, five innovation-related portfolios from Hirshleifer, Hsu, and Li (2018), Hirshleifer, Hsu, and Li (2013), Chan, Lakonishok, and Sougiannis (2001), Eberhart, Maxwell, and Siddique (2004), and Cohen, Diether, and Malloy (2013) are considered. Depending on the models employed, NMO exhibits annualized alphas ranging from 5.7% to 14.7%. When all the factors and portfolios are taken into account, NMO yields a monthly alpha of 0.01 (t=5.34), roughly equivalent to a 12% annualized alpha. The R^2 of the model is 0.58, indicating that while innovation-related portfolios account for some of the returns generated by NMO, a significant portion remains unexplained by them. The findings further validate that NMO contains valuable information not found in previously examined factors or innovation variables, as demonstrated by the consistently positive and statistically significant intercepts across all spanning regressions (Barillas and Shanken (2016)).

Next, I examine whether stock returns can be predicted by exposure to new technologies as a proxy for expected idiosyncratic risks. I run monthly Fama and MacBeth (1973) cross-sectional regressions with Newey and West (1986) autocorrelation-adjusted heteroscedasticity-robust standard errors. I investigate different model specifications where different firm characteristics are included. I find that the loading of the exposure to new technologies is 0.431% (t=2.17) in the model including market capitalization, market-to-book ratio, stock momentum, investment, operating profitability, R&D intensity, and huge increases in R&D. The result shows that the exposure to emerging technology is positively related to future stock returns on average. However, I find that the t-statistics of the coefficients of exposure to new tech are low across the regressions, highlighting the high variability of the realized returns.

The innovation activities of firms are difficult for investors to process, so previous studies often attribute higher returns associated with innovation activities to mispricing (Chan, Lakonishok, and Sougiannis (2001), Eberhart, Maxwell, and Siddique (2004), Hirshleifer, Hsu, and Li (2013), Hirshleifer, Hsu, and Li (2018), Cohen, Diether, and Malloy (2013), Leung, Evans, and Mazouz (2020), Fitzgerald et al. (2021)). I further conduct an analysis emphasizing that the findings are not solely driven by mispricing but rather by the risk-return trade-off associated with more uncertain future payoffs.

I examine the risk premiums of stocks of firms with different levels of new tech exposures. I utilize the equation that relates risk premium with risk, $E[u(W)] = u(E[W] - RP)$, where W is a payoff when a dollar is invested in a portfolio with a certain distribution, and $u(\cdot)$ is a concave utility function that exhibits decreasing marginal utility, implying that the investor is risk-averse. For random payoffs, W, a risk-averse investor knows that the expected utility of a future cash flow, $E[u(W)]$, will always be lower than the utility of receiving a certain outcome, $u(E(W))$. The risk premium (RP) is subtracted from the expected payoff to equalize them. The risk premium increases approximately in proportion to the variance of the payoff. A sufficiently high risk premium compensates for the heightened risk associated with the larger variance in the payoff.

To find the distributions of W from investing in a portfolio, I calculate return distributions using the realized one- and three-year returns of the portfolio every June. I then aggregate same-horizon returns of the portfolio from the entire sample period, dividing them into 100 equal-sized intervals to evaluate the likelihood of returns falling within these segments. This method yields return distributions for the payoffs from the portfolio, separately for one- and three-year return periods. In this way, I construct return distributions for the portfolios of firms with different levels of exposure to new technologies. Specifically, four portfolios are formed based on the level of exposure to new technologies.

I compute the risk premium from a return distribution, defined by the equation relating risk and risk premium: $E[u(W)] = u(E[W] - RP)$. A consistent pattern emerges where the risk premium for a portfolio that consists of firms with the highest exposure to new technologies is the largest. Assuming an investor with a relative risk aversion coefficient of 0.5, a one-year investment of \$1 in a portfolio with the highest exposure to new technologies requires a risk premium of 8.29 cents for the risk-averse investor. Furthermore, such an investor would demand a 21.59-cent premium for three-year investments in stocks closely related to new technologies. This pattern indicates that companies with higher exposure to new technologies display increased uncertainty in their returns, leading risk-averse investors to require a higher risk premium to hold these stocks.

I further explore the correlation between exposure to new technologies and future idiosyncratic volatility. I find that long-term idiosyncratic volatility, evaluated over the subsequent one-, two-, and three-year periods, exhibits a positive correlation with this measure, consistently observed in the cross-section and time section. The magnitude of coefficients of exposure to new technology decreases over time.

This paper contributes to the discussion on the idiosyncratic volatility puzzle that has not been resolved (Hou and Loh (2016)). I find that uncertainty related to new technology, amplifying the unpredictability of future cash flows and thereby increasing expected idiosyncratic risks, predicts positive future stock returns. One potential explanation for the famous negative relationship between past one-month idiosyncratic risks and subsequent stock returns is that high short-term past volatility signals events where uncertainties have been resolved, either in a positive or negative direction. As a result, the expected idiosyncratic risks decrease post-resolution of these uncertainties, which results in lower future returns. In other words, historical idiosyncratic volatility may represent an aggregation of resolved uncertainties or news, potentially rendering the metric somewhat outdated.

This paper is also related to several strands of literature. First, the paper is broadly connected to the asset pricing literature that relates firm characteristics to asset prices (Griffin and Lemmon (2002), Jones and Lamont (2002), Gu (2005), Zhang (2006), Fama and French (2006), Daniel and Titman (2006), Cooper, Gulen, and Schill (2008), Li (2011), Edmans (2011), Eisfeldt and Papanikolaou (2013), Fama and French (2015), Gu (2016), Lee et al. (2019), to name a few). More closely connected is the literature of empirical asset pricing that explores the implications of innovation activities of firms (Hirshleifer, Hsu, and Li (2018), Hirshleifer, Hsu, and Li (2013), Chan, Lakonishok, and Sougiannis (2001), Eberhart, Maxwell, and Siddique (2004), Cohen, Diether, and Malloy (2013), Leung, Evans, and Mazouz (2020), Fitzgerald et al. (2021)). Those studies conclude that an innovation variable can predict future returns due to the presence of mispricing. I show that there is an innovation variable (the exposure to new technologies) that is associated with risks, and the risks associated with the variable propagate to stock prices.

Lastly, I focus on a specific aspect of innovation, which is the extent to which innovation relates to emerging technologies, and I provide a new way of identifying it. There are papers that attempt to measure innovation broadly (Kelly et al. (2021), Bellstam, Bhagat, and Cookson (2021), to name a few) as well as papers zero in on a particular facet of it (for example, Kaplan and Vakili (2014), Kim and Bae (2017), Ma (2021)). I focus on the level of exposure to emerging technologies because of the uncertainty surrounding these technologies and their potential effect on stock returns. To identify new technologies, I not only utilize text data but also incorporate network data of patent publications. As a result, the deep learning model I construct employs richer information than what previous models have used to detect new technologies (Érdi et al. (2013), Furukawa et al. (2015), Breitzman and Thomas (2015), Kim and Bae (2017), Kyebambe et al. (2017), Lee et al. (2018), Zhou et al. (2020), Arts, Hou, and Gomez (2021), Zhou et al. (2021), Choi, Park, and Lee (2021)). More specifically, I analyze not only the focal invention's text data (abstract, claim, and description) but also the text information of neighboring patents related to the focal invention. The deep learning model I employ for the paper is trained in a way that the model effectively summarizes the vast amount of information and produces an embedding that describes the focal invention.

The remainder of the paper is structured as follows: Section 1 delves into the data utilized in this study. Section 2 outlines the process of training a deep learning model to summarize information on patent publications. Section 3 presents a method for identifying new technology clusters and their characteristics. Section 4 carries out a portfolio analysis. Section 5 offers potential explanations for the performance of the NMO portfolio. Finally, Section 6 provides concluding remarks.

2.1 Data

Patent data are collected from Google Patent.[13] I start with all publicized US patents and patent applications since 1976 irrespective of whether or not they are withdrawn, expired, pending, or currently active. A patent publication refers to an invention whose information is released publicly by the USPTO either because patent protection is granted or because the information of patent application has become available due to the American Investors Protection Act of 1999.[14] I keep utility patent data excluding design, and plant patents.

On top of the U.S. patent publications issued after 1976, I collect information on all U.S. patents granted before 1976 and international patent publications that are one or two hops away from a publication in a citation network.[15] To be more specific, let us say a publication A is a U.S. patent publication issued after 1976. A patent publication B is one hop away from A if A cites B or B cites A. The information of B is gathered. A patent publication C is two hops away from A if C cites a patent publication that is one hop away from A, for instance, B, but C itself is not one hop way from A. The information of C is collected.

The sample selection criterion of having to be located within two hops from the U.S. patent publications issued after 1976 is set because a graph neural network (GNN) model that produces embeddings (a vector representation) of a patent publication uses the information of their neighborhood patents. More specifically, I build a one-layer GNN model that generates a vector for each U.S. patent publication issued after 1976 with an aim to use the vector to measure the similarity between an invention and other inventions. The GNN model uses not only the information of an invention of interest itself but also all the information of adjacent publications linked by citation relationships. The sample selection enables me to avoid any subgraph

[13] I respected the policy specified in https://patents.google.com/robots.txt to gather data by web scraping. Specifically, I used the path '/patent/'.

[14] All patent applications' information should be published 18 months after the application date thanks to the enactment of the American Investors Protection Act of 1999.

[15] The World Intellectual Property Organization (WIPO) publishes an application as 18 months have elapsed since the initial filing date.

that is truncated due to a lack of data when training the model.

In most cases, an innovative work has more than one patent application or patents granted in countries in which inventors wish to have patent protection, regardless of whether it originated in the US or other countries. I regard them as the same invention. For example, the publications 'Property management on a smartphone (AU2016202022B2)' and 'Property management on a smartphone (CA3151541A1)' are treated as the same invention because they are the same inventions published in different countries. Moreover, an invention can have multiple publications in the U.S. when it is a continuation of a prior work or it was published due to the American Investors Protection Act of 1999 and then later granted patent protection. Mechanically, I regard the multiple publications of an invention as the same invention. For example, the publications 'US20070263207A1' (published due to the American Investors Protection Act of 1999) and 'US7894055B2' (published when a patent is granted) are the same invention. For instance, when the publication 'US20150090883A1' cited 'US20070263207A1', and 'US8513981B2' cited 'US7894055B2', I conclude that both 'US20150090883A1' and 'US8513981B2' cited the same invention. There are other cases where two publications indicate the same technology, and I treat them as the same invention.[16] After having unique inventions, I link them with the earliest publication date for each invention. The resulting dataset is a set of unique inventions and the earliest publication date of each invention.

I gather CIK and historical company names registered with the SEC from the SEC bulk data set. Subsidiary names are obtained by cleaning the subsidiary data provided by WRDS between 1995 and 2019. To gather subsidiary data prior to 1995 and after 2019, I collect Exhibit 21 of Annual Report 10-K. Whenever there is no subsidiary information, I infer that the subsidiaries of a firm are the same as those in the previous year. To match assignees of a patent or application to US companies, I eliminate company abbreviations and find the exact match. A more detailed matching process and results are provided in Appendix A.

The sample consists of firms in the intersection of Compustat, Center for Research in Security Prices (CRSP), and the patent data set. Following Fama and French (1993), all U.S. common shares (CRSP 'shrcd' in (10, 11)) trading on NYSE, AMEX, and Nasdaq are included. Financial companies with SIC between 6000 and 6999 are

[16]Technically, I identify unique inventions by the earliest priority number whenever possible. I eliminate any space in a priority number string and made adjustments such as replacing 'JPJP' with 'JP' whenever I find inconsistent representation of priority numbers, which were very rare. If the earliest priority number does not exist, I use the earliest publication number to identify unique inventions.

excluded. Following Hirshleifer, Hsu, and Li (2013), firms are included if they have been listed on Compustat for 2 years before to mitigate backfilling bias.

The one-month Treasury bill rate, the market excess return, SMB, HML, robust minus weak returns (RMW), and conservative minus aggressive return (CMA) are obtained from Kenneth French's website. Accounting variables are collected from Compustat Fundamentals.

2.2 Invention Embeddings

In this section, I explain how I construct features of an invention that will be used to measure similarity between U.S. patent inventions issued after 1976. I use a graph neural network (GNN) model where I not only use an invention's information but also all the information of the patent inventions relating to the focal invention. It is because the information surrounding the focal invention can be crucial in inferring the characteristics of the focal invention (Jaffe, Trajtenberg, and Fogarty (2000), Acemoglu, Akcigit, and Kerr (2016), Watzinger and Schnitzer (2019)). The advantage of using GNN models is that I can tap into the idea that two inventions must be similar if they share many backward and forward citations. The weight or importance of a neighborhood invention in computing the features of the focal invention will be empirically determined by a deep learning method.

In the first step, I collect all inventions that are one or two hops away from a U.S. patent invention issued after 1976 to feed into a one-layer GNN model. The model is one layer in the sense that, to generate a vector representation for a patent invention of interest, it inputs neighborhood inventions (and the invention itself) that are directly connected to the invention through citations. On top of inventions in a direct citation network, the model requires inventions that are two hops away from a U.S. patent invention because the loss function requires the embeddings of one-hop neighborhood inventions computed by the model. To understand this, see Figure 1. The white nodes are U.S. patent inventions issued after 1976 and the gray nodes are those not. The circles with a thick outline are collected while the dashed circles are not collected. The gray lines are citation relations depicted irrespective of the direction of citations. Let us say the node labeled v is an invention of interest. The goal is to train the one-layer GNN model so that it produces an embedding of v that is similar to the embeddings of $u_1, u_2, \ldots, u_D$, where D is the degree of the node v, because they are closely related. This necessitates the embeddings of $u_1, u_2, \ldots, u_D$ generated by the model, which requires the information of neighborhood inventions of all $u_1, u_2, \ldots, u_D$. For example, the node u_1 is directly connected to the node v

while s_1 is not. s_1 is not used to generate the embedding of v but will be used to produce the embedding of u_1. Both u_1 and s_1 are collected because they are within two hops from the node v.

To produce raw features that will be fed into the model, I draw on text and network data. For each observation, I have the title and machine-extracted keywords. I refine the text data so that I form a sentence that describes the invention. Specifically, I create a sentence "The patent's title is '*title*', which is about '*keyword* 1', '*keyword* 2', ..., '*keyword K*'.", where the italic words are actual title and keywords of an invention. I cap the number of keywords included in a sentence to thirty which are sorted by the count in the invention text. In addition to text, I employ the citation network of an observation. I do not distinguish between backward citation and forward citation as long as they are made before the year of portfolio rebalancing.

There are a number of GNN models to choose from. I selected the Graph Attention Network (GAT, Graph ATtention network) model (Veličković et al. (2017), Brody, Alon, and Yahav (2021)) because the model allows me to empirically determine which neighborhood patent inventions the model should pay more attention to when computing the features of focal patent. I add a Bidirectional Encoder Representations from Transformers (BERT, Devlin et al. (2018)) layer and a Long Short-Term Memory (LSTM, Hochreiter and Schmidhuber (1997)) layer before the GAT model to process text data before utilizing network features. The architecture of the model is depicted in Figure 2. The caption of Figure 2 provides more detailed information on the GNN model.

The model learns parameters in an unsupervised way by minimizing the following loss function which is based on Hamilton, Ying, and Leskovec (2017).

$$loss(z_v) = -\frac{1}{D}\sum_{u \in U(v)} log(\sigma(z_u^T z_v) + \epsilon) - \frac{1}{D}\sum_{w \in W(v)} log(\sigma(-z_w^T z_v) + \epsilon) \qquad (6)$$

z_v is a vector of an invention v produced by the model. z_u is a vector of an invention that is one hop away from v, also produced by the model. $U(v)$ is a set of vectors of v's neighborhood inventions. z_w is a vector of an invention in a negative sample $W(v)$ of v produced by the model. Simply, I choose a negative sample from all other U.S. inventions issued after 1976 and before a hypothetical rebalancing day with the sample size being $2 \times D$, where D is the degree of v. Here, the number of negative sample is two and each negative sample is of size D. The choice of the parameters of the negative samples are based on Mikolov et al. (2013). ϵ is added to avoid $log(0)$ and σ is a sigmoid function. The idea behind the loss function is that we want a large inner product of z_u and z_v ($z_u \in U(v)$) because they are related, but a small inner product of z_w and z_v ($z_w \in W(v)$) because they are likely unrelated.

It is too costly to input all 30 millions collected inventions to train the model. To effectively train the model without incurring excessive costs, a random sample of 2% of U.S. inventions from the past 15 years is chosen, with a focus on those issued after 1976. Additionally, all one and two-hop neighborhood inventions are included. For negative samples, a random selection from U.S. inventions issued post-1976 is made, with the sample size being double the number of neighborhood inventions. All their neighborhood inventions are included to compute the embeddings of the negative samples.

The training set consists of a U.S. invention issued after 1976, its one and two-hop neighborhood inventions, and a negative sample that's twice the size of the central node's degree, and their neighborhood inventions.

To maintain consistency with rebalancing portfolios based on public information, the features of inventions are computed using only available data at the time of rebalancing. Ideally, the model should be retrained every year using the information available up to that point. However, in the interest of time, the model is trained only at the end of June every five years (1981, 1986, 1991, 1996, 2001, 2006, 2011, 2016) using the past fifteen-year inventions. Then, an embedding of an invention is computed with the latest model that is hypothetically available by the time of rebalancing. This way, I avoid look-ahead bias.

2.3 The Exposure to New Technology

In this section, I describe the construction of a firm-level measure for exposure to new technologies. At the end of each year t, I perform clustering analysis using patents published in the past ten years. The input for this analysis is 256-number vectors generated by the deep learning model explained in the previous section. Each vector represents an invention and serves as condensed information that summarizes the text data of the corresponding invention (claim, abstract, description), as well as the text data of its neighboring patents (publications that appear in the invention's references). I compute the cubic clustering criterion to find the optimal number of clusters. At the end of each year t, clusters are sorted based on the proportion of patents publicized in the recent three years ($t-3$ to t). In an independent sort, clusters are sorted based on the proportion of patents publicized in the earliest three years ($t-10$ to $t-7$). The new technology clusters are defined as the intersection of the top thirty percent from the first sort and the bottom thirty percent from the last sort.

The rationale behind this approach is twofold. First, according to Rotolo, Hicks, and Martin (2015), one characteristic of emerging technologies is their relatively fast growth. This idea is reflected in the classification process that selects clusters with more recently published inventions while eliminating those with many older inventions. Second, another attribute of emerging technologies is coherence. This concept is captured by the clustering analysis method. At the end of each year t, I generate the embeddings of all U.S. inventions in the past ten years using the same deep learning model so that each dimension of the embeddings indicates the same information. The time span of ten years and the volume of information used to detect new tech clusters allow me to identify coherent clusters at any given point during the sample period.

For inventions publicly reported in year t, I categorize them as either related to new technologies or not, based on the clusters they belong to. Essentially, each invention is identified as a new technology or not just once, based on the year it first becomes publicly available and the clusters formed that year. When conducting a clustering analysis for subsequent years, I disregard the information. I repeat this process annually, classifying all inventions. The purpose of classifying inventions only once is to capture a company's activities in a specific year. When companies develop inventions related to new technologies during a specific time period, it demonstrates their engagement in emerging technologies, irrespective of the point of evaluation in the timeline.

A variable is created using the information. $NewTech$ is an invention-level variable that takes one if an invention belongs to a cluster that is classified as new technologies at the time the invention was revealed to the public for the first time. The summary statistics of $NewTech$ is available in Table 1. The mean value of $NewTech$ is 0.301 meaning that for all new inventions about 30% are classified as relating to new technologies under the scheme of identifying new tech clusters.

I investigate the attribute of the variable $NewTech$ in Table 2 where I present results from invention-level regressions. I select the following characteristics of inventions. The variable $cite\_less2y$ is the fraction of patent references (citations) that are publicized less than two years before the information of the invention is publicly available. It measures to what extent an invention draws on recent innovative work. The variable $cite\_age\_std$ measures the variation in the age of patent references an invention is built on. $cite\_patentN$ is the number of patent references, while $cite\_nonpatentN$ is the number of non-patent references such as academic papers or research papers by companies. $cite\_countryN$ measures the number of countries of patent references. $cite\_categoryN$ measures the number of technological categories of patent references where categories are defined by the classifications of the Cooperative Patent Classification. $public$ is an indicator variable that takes one if the invention's original assignee is a public company or a subsidiary of a public company, zero otherwise. $publlicShare$ is the fraction of patent references whose original assignee is a public company or a subsidiary of a public company.

The first four regressions in Table 2 show that inventions that cite younger patent references are more likely to belong to a new technology cluster. The finding shows that inventions relating to new technologies have a scientific base that is younger than other inventions in terms of the age of patent references. Next, the diversity in the reference age is positively related to new technologies. The finding is seemingly at odds with the first finding. I interpret it as an invention of new technologies drawing on a relatively more diverse set of knowledge along with other dimensions of diversity such as technological categories. Next, the number of patent references is negatively related to new technologies, implying that new tech inventions rely less on previous patents. On the other hand, new tech inventions are more likely to cite more non-patent references. This signifies that new tech inventions actively utilize knowledge outside the existing patent universe. The number of countries of patent citations also explains new technologies but becomes statistically insignificant once category-fixed effects are included. The number of technological categories of patent references is positively related to new technologies. This finding is consistent with the idea that new technologies are created by connecting different ideas from various areas. Also, public firms are more likely to be engaging in developing new technologies, which implies that exploring new tech requires huge resources. Similarly, the fraction of assignees of patent references that are public firms is positively related to new technologies.

In the last four columns, I add variables that are correlated with some variables in the first four regressions that nonetheless may convey different information. $cite\_age$ is the median age of patent references. Its correlation with $cite\_less2y$ is -0.74. $cite\_countryHHI$ is a measure of the concentration of countries of patent references which is calculated by squaring the share of each country and then summing the resulting numbers. Its correlation with $cite\_countryN$ is -0.74. $cite\_categoryHHI$ is a measure of the concentration of categories of patent references which is calculated by squaring the share of a classification of Cooperative Patent Classification and then summing the resulting numbers. Its correlation with $cite\_categoryN$ is -

0.67. $cite\_firmHHI$ is a measure of the concentration of firms of patent references which is computed by squaring the share of each firm and then summing the resulting numbers. Its correlation with $publlicShare$ is 0.08.

The additional variables that convey incremental information are $cite\_age$ and $cite\_firmHHI$. The coefficients of $cite\_age$ show that inventions associated with new technologies are more likely to cite younger patent references. The coefficients of $cite\_firmHHI$ show that when an invention cites patents of various companies, it is more likely related to new technologies. Overall, the regression results implemented at the invention level show that the relation between invention characteristics and new technology clusters is consistent with our intuition about what constitutes new technologies.

Utilizing invention data, I calculate firm-level exposure to emerging technologies. This exposure is determined using the information on new inventions from the previous three years, specifically, those published for the first time during this period. Exposure to new technologies is defined as the proportion of new inventions related to emerging technologies among all newly introduced inventions within the past three years. This metric is denoted as $NewTechExposure$.

Table 1 presents the summary statistics for this variable, with a mean of 0.298 and a standard deviation of 0.349, indicating considerable variation in its values. The 25th percentile of the variable is zero, suggesting that even patent-producing firms may not consistently engage in R&D activities pertaining to new technologies. The analysis was replicated using a measure of exposure based on new inventions from the previous five years, yielding similar conclusions. In fact, the correlation between the two exposure measures is 95%.

In Table 3, I report the level of exposure to new technologies by industry. The table shows the summary statistics of the level of exposure to new technologies for industries defined by Fama & French twelve-industry classification. Financial firms are excluded from the sample (code 11).

The industries with high exposure to emerging technologies are Business Equipment (code 6), Telephone and Television Transmission (code 7), and Healthcare, Medical Equipment, and Drugs (code 10). However, the dispersion of the variable is higher for those industries, too. The significant variation within each sector is noteworthy, as it suggests that taking a long or short position in a stock based on its exposure to new technologies does not equate to adopting a specific position in a stock purely based on its industry.

To understand the composition of the sample, I compute the fraction of firms in the CRSP database that have information on the exposure to new technologies in the last column. I find more than half of the firms in Consumer Durables (code 2), Manufacturing (code 3), and Healthcare, Medical Equipment, and Drugs (code 10) have patent publications data needed to compute the exposure to new technologies. However, these sectors do not always exhibit high levels of exposure to new technologies, indicating that the extent of patenting activities and the degree of engagement with new technologies are separate concepts.

I further examine individual stocks. Figure 3 shows the time-varying exposure of stocks of Meta Platforms, Amazon, Apple, Microsoft, and Alphabet (MAMAA stocks). The plots for Amazon, Alphabet, and Meta stocks commence at specific points during the sample period, coinciding with their respective initial public offerings. Consistent with the fact that MAMAA firms are known for their influence on the technology landscape, the exposure of the companies to new technologies is higher than both the sample mean of 0.30 and the median value of 0.15.

Several distinct patterns emerge from these plots. Firstly, the level of exposure remains relatively stable over time, with no substantial fluctuations, indicating that the technologies employed by these firms evolve in a consistent manner. Secondly, the exposure to emerging technologies for four of the companies—excluding Meta—appears to have diminished over time. Thirdly, the exposure of all five stocks tends to follow a similar trajectory, potentially due to the interconnected nature of their respective technologies. Lastly, in accordance with expectations, Meta, the most recent entrant in the group, exhibits a higher level of exposure to new technologies compared to its counterparts.

Next, I construct other innovation-related variables that are known to explain stock returns. $InnOrig$ is a measure of the innovative originality of a firm introduced by Hirshleifer, Hsu, and Li (2018). To compute the variable, an individual invention's citation diversity is measured. It is the log of (one plus) the number of unique technological classes assigned to the patents cited by the invention. I use classes defined by the Cooperative Patent Classification. Then, a firm's $InnOrig$ is calculated as the average citation diversity across all inventions over the past five years.

While the measure might be connected to valuation uncertainty, it is important to note that it is distinct from exposure to emerging technologies. Specifically, the variable captures the range of technology a patent draws on, as well as the capacity of a firm's managers and scientists to effectively combine different technologies in an innovative manner. As a result, the measure is primarily linked to the degree of information that investors must digest, which can be challenging and, in turn, contribute to mispricing.

$efficiency$ is a measure of a firm's ability to efficiently generate innovative work defined as the number of patents over R&D capital introduced by Hirshleifer, Hsu, and Li (2013). The measure is defined as the ratio of firm i's patents granted in year t to its R&D capital. R&D capital is measured using the R&D expenses incurred by the firm in the years between $t-6$ and $t-2$. It is the 5-year cumulative R&D expenses assuming an annual depreciation rate of 20% as in Chan, Lakonishok, and Sougiannis (2001). When I construct this particular variable, I follow Hirshleifer, Hsu, and Li (2013) to set missing R&D to zero.

According to Hirshleifer et al. (2013), limited attention among investors drives the positive association between efficiency and stock returns. Specifically, innovative efficiency is an asset that is associated with future operating performance. Nonetheless, it may not be immediately salient or easy to process, leading to mispricing.

$R\&Dintensity$ is a measure of R&D intensity defined as R&D expenditure over contemporaneous sales (Chan, Lakonishok, and Sougiannis (2001)). R&D-intensive firms often lack tangible assets and depend on untested technologies, making their prospects highly unpredictable. Therefore, stock prices do not fully value R&D, which could be an invaluable intangible asset.

$huge\Delta R\&D$ is the significantly positive abnormal increase in R&D expenditures defined by Eberhart, Maxwell, and Siddique (2004). To construct this variable, I begin with firm-year observations where $\text{R\&D}_t$ is greater than $\text{R\&D}_{t-1}$ in any year t between 1980 and 2019. Unexpected R&D increases are defined as cases where $\text{R\&D}_t/\text{Total Assets}_t$ is greater than $\text{R\&D}_{t-1}/\text{Total Assets}_{t-1}$. Then, firms with $\text{R\&D}_t/\text{Sales}_t$ greater than 0.05 are chosen because it shows the economic importance of R&D to the firm. Finally, a huge increase in R&D satisfies $\text{R\&D}_t/\text{R\&D}_{t-1} >$ 1.05 and $(\text{R\&D}_{t-1}/\text{Total Assets}_{t-1})/(\text{R\&D}_{t-1}/\text{Total Assets}_{t-1})$ is greater than 1.05. $huge\Delta R\&D$ is an indicator variable for firm-year observations where the aforementioned conditions are all satisfied. Otherwise, the indicator variable is set to zero.

Eberhart, Maxwell, and Siddique (2004) find positive abnormal stock returns observed after R&D increases occurred from 1951 to 2001. This suggests that increases in R&D are advantageous for companies and that the market may take time to fully appreciate the magnitude of this benefit.

Finally, $InnoAbility$ is a measure of a firm's ability to turn R&D expenditures into future sales introduced by Cohen, Diether, and Malloy (2013). It is computed from the estimates of the following rolling firm-by-firm regressions of firm-level sales growth on the log of lagged R&D scaled by sales.

$$log(\frac{Sales_{i,t}}{Sales_{i,t-1}}) = \beta_0 + \beta_\tau log(1 + \frac{R\&D_{i,t-\tau}}{Sales_{i,t-\tau}}) + \epsilon_{i,t} \quad \tau \in \{1,2,3,4,5\} \tag{7}$$

Then, the five estimates of $\hat{\beta_\tau}$, $\tau \in \{1,2,3,4,5\}$, at a given t are averaged to construct the variable $InnoAbility$. There are several restrictions used by Cohen, Diether, and Malloy (2013). In estimating the coefficients, for each regression, I use eight years of past data where at least six R&D observations are non-missing. Also, half of the non-missing R&D observations should be positive. If the two conditions are not met, the slope coefficients are set to missing values.

Similar to $efficiency$, a firm's ability to transform R&D to sales is an important piece of information on the firm's future prospects. However, investors fail to fully incorporate the information, leading to a positive relationship between $InnoAbility$ and future stock returns.

Other accounting variables are included in some model specifications. $size$ is the logarithm of total assets. $cash$ is cash and cash equivalent divided by total assets. ml is market leverage. mb is market-to-book ratio. All accounting and innovation-related variables are winsorized at the 1% and 99% levels.

Table 4 shows the relationship between a firm's exposure to new technologies and other innovation variables. First, a firm's innovation originality ($InnOrig$) has a positive coefficient in the model (1). The positive relationship could be due to the uncertainty associated with innovation originality or the likelihood that firms that exhibit more originality tend to produce inventions relating to new technologies as suggested by the preceding patent-level analysis. However, upon incorporating accounting variables (as seen in column (6)), the coefficient loses its statistical significance. This indicates that exposure to emerging technologies and innovation originality convey substantially distinct information.

A firm's capacity for efficiently generating innovative output ($efficiency$) exhibits a negative and statistically significant correlation with $NewTechExposure$. This suggests that the number of patents produced per unit of R&D capital does not directly correspond to the development of new technologies. This observation aligns with the intuitive understanding that investments in emerging technologies often require an extended period before yielding tangible results. As companies allocate resources towards exploring and developing novel technologies, the immediate efficiency of their R&D efforts may be reduced.

The relationship between a company's R&D expenditures relative to sales and

$NewTechExposure$ is not straightforward. The coefficients in model (3) and model (8) display opposite signs, indicating an inconclusive connection between these variables. It is possible that the significance of R&D investments within a firm may not provide adequate insights into whether the company is actively exploring new technologies or focusing on innovations closely tied to existing technologies.

This lack of clarity could stem from various factors, such as differences in industry dynamics, company size, or the stage of technological development. For example, smaller companies or those in rapidly evolving industries may allocate a higher proportion of their resources to R&D, while larger firms or those in more mature industries may have lower R&D intensity. Additionally, some firms might focus on incremental improvements of existing technologies, while others may prioritize groundbreaking innovations, both of which can influence the relationship between $R\&Dintensity$ and $NewTechExposure$.

A substantial increase in R&D ($huge\Delta R\&D$), on the other hand, demonstrates a positive correlation with $NewTechExposure$. One possible explanation for this relationship is that firms with high exposure to new technologies may require consistent, additional resources to continue delving into these emerging areas. The rapidly evolving nature of new technologies demands continuous investment in R&D to maintain a competitive edge and capitalize on potential breakthroughs.

Lastly, a firm's capacity to convert R&D investments into sales ($InnoAbility$) demonstrates a negative relationship with the new tech exposure, albeit with limited statistical significance. This finding echoes the connection observed between the exposure to new tech and efficiency. As a firm ventures into new technological areas, its ability to transform innovation into sales may be hindered, primarily due to the indefinite time it takes to fully understand, develop, and commercialize these emerging technologies. Additionally, the costs associated with developing and commercializing new technologies can be substantial, potentially affecting a firm's short-term profitability and financial performance. As a result, firms that actively explore new technologies may experience a temporary decline in their innovation-to-sales conversion ratio, reflecting the complex and time-consuming nature of pioneering technological advancements.

The papers introducing the five innovation-related variables as relating to mispricing posit that information related to a company's R&D activities is complex to process. This complexity implies that the data concerning these variables may not be immediately incorporated into stock prices. The correlation between these innovation variables and exposure to new technologies indicates that the metric for uncertain future cash flows also encompasses the potential for mispricing. Notably, the distinct aspect of the new technology exposure metric, compared to the five innovation variables, is its explicit association with idiosyncratic risks. In contrast, the other variables do not have a direct link to the variability in future cash flows.

In summary, the analysis of the relationship between exposure to new technologies and other innovation variables, which are known to forecast positive future stock returns, reveals that $NewTechExposure$ displays traits in line with prevailing assumptions about the nature of emerging technologies. Also, $NewTechExposure$ contains distinct information from other innovation variables.

2.4 The Exposure to Emerging Technology and Stock Returns

In this section, I carry out a portfolio analysis to determine if firms with high exposure to emerging technologies yield higher returns. In the first subsection, I conduct a sort analysis. In the next section, I compute alphas of portfolios that have different levels of exposure to new technologies. The return predictive power of the exposure to new technologies is documented in the last subsection.

2.4.1 Monthly Returns of New Tech Exposure Portfolios

In this subsection, I examine the monthly returns of portfolios according to their exposure to new technologies. Each year, portfolios are formed based on their exposure to new technology. Notably, over 25% of the firms display zero $NewTechExposure$, despite having applied for patent grants. This implies that these firms may have sought patents in areas not classified as new technologies or that their patent applications do not significantly impact their overall exposure to emerging technologies. As a result, I begin by classifying firms into those with zero exposure ('No' exposure group) and those with positive exposure at the end of June of each year. For firms with positive exposure, I sort them by their $NewTechExposure$ at the end of June of each year and establish three portfolios ('Low', 'Middle', and 'High') using the 30th and 70th percentiles of exposure as cutoff points. Portfolios are assembled at the end of each June, and these portfolios are maintained for the subsequent twelve months (from July of year t to June of year $t+1$), determining their value-weighted monthly returns.

I further consider the effect of size and conduct a double-sort analysis. Following Fama and French (1996), at the end of June of each year t, stocks are independently sorted into two groups ('Small' or 'Big') based on their June market capitalization relative to the median market capitalization of NYSE stocks. Size-adjusted new tech exposure portfolios are created by computing the average of double-sorted portfolios. 'Zero' portfolio is (No/Small +No/Big)/2. 'Old' portfolio is (Low/Small +Low/Big)/2. 'Mid' portfolio is (Middle/Small +Middle/Big)/2. Similarly, 'New' portfolio is (High/Small +High/Big)/2.

Table 5 presents the performance of new technology exposure portfolios. t-statistics are displayed in square brackets, and all returns are expressed in decimal form. Note that the sample comprises firms actively involved in patenting. The absence of negative returns in these portfolios suggests that firms consistently engaged in patent creation tend to see positive future returns (Bedford et al. (2021)).

A notable pattern emerges, with the 'High' portfolio consistently outperforming all other exposure groups. In the single-sorted portfolio, the average monthly return of the 'High' portfolio is 0.0171 ($t = 6.50$). In the double-sorted portfolio analysis, the 'High/Small' portfolio generates 0.0337 ($t = 6.89$), while the 'High/Big' portfolio yields 0.0154 ($t = 6.08$). Moreover, the 'Middle' portfolio outperforms both the 'No' and 'Low' portfolios in both single and double-sort analyses. Turning to size-adjusted new tech exposure portfolios, I find that the new minus zero portfolio and the new minus old portfolio both generate positive average monthly returns. For the new minus old portfolio, the average monthly return is 0.006 ($t = 2.99$). The number corresponds to 7.4% annualized returns ($= (1.006)^{12} - 1$).

In both single- and double-sort analyses, a clear distinction between the 'No' and 'Low' portfolios is not apparent. Also, NMZ and NMO do not show a stark difference, possibly due to shared characteristics. In other words, firms with zero exposure to new technologies may not be significantly different from those with minimal exposure. However, as will be demonstrated in Section 5, stocks with zero exposure to new technologies could be riskier than those with at least some exposure to new technologies. Consequently, I focus on the NMO portfolio, which highlights the contrasting uncertainties associated with these two types of stocks ('New' and 'Old'). The overall findings in Table 5 indicate that firms with greater exposure to new technologies generate higher returns.

2.4.2 The Relative Performance of New Tech Exposure Portfolios

The performance of a new minus old portfolio may be explained by other risk or mispricing factors such as size or innovation-related portfolio returns that have already been explored in the literature. In this subsection, I examine the performance of the NMO portfolio, which is associated with uncertainty in emerging technologies, in comparison to widely recognized risk factors and other innovation-related portfolios.

The alphas and R^2 of new tech exposure portfolios are reported in Table 6. I first consider the four-factor (market, size, value, momentum factors) model (4F) (Carhart (1997)). The alpha of NMO is 0.0063 ($t = 3.41$, $R^2 = 0.22$). The number implies a 7.83% annualized return. After including the robust-minus-weak (RMW) factor and the conservative-minus-aggressive (CMA) factor from Fama and French (2015), the magnitude of the alpha of NMO increases dramatically. The alpha of NMO is 0.0096 ($t = 5.29$, $R^2 = 0.30$).

I also examine innovation-related portfolios' returns introduced by previous studies and augment them to the 4F plus RMW+CMA model. The following portfolios are considered: ORIG is the monthly returns of a portfolio that longs firms with high originality and shorts firms with low originality, as constructed by Hirshleifer, Hsu, and Li (2018). EFF is a portfolio that longs firms with high innovation efficiency and shorts firms with low efficiency, as developed by Hirshleifer, Hsu, and Li (2013). INT is a portfolio that longs firms with high R&D intensity and shorts firms with low R&D intensity, as measured by Chan, Lakonishok, and Sougiannis (2001). INC is a strategy that longs firms experiencing significant increases in R&D expenditure, as identified by Eberhart, Maxwell, and Siddique (2004), and shorts the 3-month treasury bill. ABI is a portfolio that longs firms with high R&D expenditure and high ability to convert it into sales, while shorting firms with high R&D expenditure and low ability, as measured by Cohen, Diether, and Malloy (2013).

In all the model where a single innovation-related portfolio is added to the 4F+ RWM +CMA model, the alphas of NMO is positive and statistically significant. The lowest number is 0.0047 ($t = 2.45$, $R^2 = 0.36$) in the 4F+RMW+CMA +INC model, which corresponds to 5.7% annualized returns. The largest number is found in the 4F+RMW +CMA +EFF model. The alpha of NMO is 0.0115 ($t = 6.85$, $R^2 = 0.51$), which is 14.7% annualized return. In the last column, I include all five innovation-related portfolios and still find economically large and statistically significant alpha (0.01, $t = 5.34$, $R^2 = 0.58$).

The regression presented in the last column also acts as a spanning regression (Barillas and Shanken (2016)). The most comprehensive model, which includes all innovation portfolios, still leaves substantial alphas on NMO. These statistically significant alphas suggest that NMO offers valuable information about expected returns when compared to the model, or in other words, an asset pricing model augmented with NMO outperforms the model containing previous innovation portfolios. Overall, the results demonstrate that NMO's performance is not subsumed by other innovation variables, and most importantly, the returns generated by NMO are economically significant.

2.4.3 *The Predictive Power of New Tech Exposure*

This subsection investigates the ability of exposure to new technologies to predict the cross-section of stock returns. I conduct monthly Fama and MacBeth (1973) regressions from July 1981 to June 2019 with monthly excess returns. The regressions are designed such that all independent variables, including $NewTechExposure$, are available prior to the observation of monthly returns. To achieve this, I use beginning-of-year accounting and innovation variables and associate them with monthly returns starting from July of the current year up to June of the following year. I use Newey and West (1986) heteroscedasticity-robust and autocorrelation-adjusted standard errors.

For other covariates, I use $MktCap$ which is the log of market capitalization, market-to-book ratio (mb), and $momentum$ defined as past six-month share price run-ups measured at the end of June. Operating profitability (op) is revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by book equity. Investment (inv) is the change in total assets from the fiscal year ending in year $t-2$ to the fiscal year ending in $t-1$, divided by $t-2$ total assets. Also, I consider two innovation variables $R\&Dintensity$ and $huge\Delta R\&D$. There are only a few observations that overlap between the accounting variables and other innovation variables each month. The significant loss of observations occurs when including $InnOrg$, $efficiency$, and $InnoAbility$. Therefore, I excluded the three variables from the analysis. All independent variables are winsorized at the 1% and 99% levels and standardized.

The results are presented in Table 7. The coefficients are reported in percentages and corresponding t-statistics are provided. The column (1) shows that the slope on $NewTechExposure$ is 0.376% (t = 1.87). The coefficients of $MktCap$ and mb, 0.332% (t = 7.47) and -0.102% (t = -3.74) are consistent with the literature. In column (2), the loading of $NewTechExposure$ is 0.420% (t = 2.07). The signs of other accounting variables are as expected except for op and $momentum$. op has a positive coefficient without the presence of other variables but shows a negative coefficient in column (2). In column (3), the coefficient of $momentum$ is negative and statistically significant throughout the model specifications, which might be a result of the sample containing firms that are actively engaging in patenting activities. The effect of $momentum$ might be more prevalent in a larger set of companies.

In column (4), $R\&Dintensity$ and $huge\Delta R\&D$ are added. $huge\Delta R\&D$ has a coefficient (0.329%, t = 3.08) whose magnitude is smaller than the loading of $NewTechExposure$ (0.441%, t = 2.15). The result shows that $NewTechExposure$ retains its predictive power when the size of R&D investments of a firm is considered. The coefficient of $R\&Dintensity$ is 0.06% (t = 1.91). When including different combinations of other accounting variables, the coefficients of $R\&D$-related variables do not change their sign, while the magnitude or statistical significance of them vary slightly. In the last column, where all variables are considered, the coefficient of $NewTechExposure$ is 0.431% (t=2.17). The result shows that $NewTechExposure$ can explain cross-sectional stock returns even when variables measuring aspects of R&D are considered.

Overall findings show that exposure to new technologies can explain cross-sectional stock returns. In particular, the economic magnitude is not trivial. Specifically, a one standard deviation increase in $NewTechExposure$ (column 7) leads to a monthly excess return increase of 0.151% on average. However, relative to other accounting or innovation variables, the statistical significance is limited. This implies that the variability of future stock returns is also large. The next question is whether the predictive power of $NewTechExposure$ is driven by risks associated with new technologies.

2.5 Exposure to New Technologies and Risks

2.5.1 The Analysis on Risk Premium

In this section, I investigate the risk premiums extracted from power utility functions in the context of investing in stocks of firms with different levels of exposure to new technologies. The risk premium (RP) is defined by the following equation: $E[u(W)] = u(E[W] - RP)$. $E[\cdot]$ is the expectation associated with a certain return distribution and $u(\cdot)$ is a utility function. For a constant γ with $0 < \gamma < 1$, u(W) = $\frac{1}{1-\gamma}W^{1-\gamma}$. In the context of power utility functions, the Arrow–Pratt measure of Relative Risk Aversion (RRA) under power utility is γ. When $\gamma > 0$, the utility function is concave, and the individual is considered risk-averse. The higher the value of γ, the more risk-averse the individual. W is defined as a dollar amount of wealth when a dollar is invested in a stock of a firm with a certain level of exposure to new technologies.

In order to find the distribution of wealth (W), I construct an implied return distribution based on realized return distributions of new tech exposure portfolios. Specifically, for each portfolio, I create 100 bins of one- and three-year cumulative returns respectively, where each bin has the same length of intervals. I choose the midpoint of the intervals as a representative return of the bin. The probability of returns falling into each bin is calculated by dividing the number of returns within a specific bin by the total number of returns across all bins.

I construct the return distributions for four groups of firms that are categorized based on firms' exposure to new technologies. Firms classified as the 'Zero' group have a $NewTechExposure$ value of zero, indicating no exposure to new technologies. For firms with positive exposure to new technologies, three distinct groups are created: 'Old', 'Mid', and 'New'. These groups are formed based on the 30th and 70th percentiles of $NewTechExposure$, allowing for a clear categorization of the firms' involvement with new technologies.

The return and corresponding wealth distributions are provided in Appendix E. The return distribution exhibits a degree of coarseness, as the first and last bins possess a substantially greater number of observations, attributable to the winsorization of returns at the 1% and 99% levels. Furthermore, the computation of the distribution does not take into account other risk factors, such as firm size, which are incorporated when constructing portfolios. Moreover, the return distribution does not account for effects arising from various time periods. It should also be noted that the return distributions exclusively consider firms with either one or three-year cumulative returns. Therefore, those firms that went bankrupt in one or three years are not included.

Table 8 presents the risk premia derived from power utility functions under various return distribution scenarios. Firstly and most importantly, there is a discernible pattern wherein the risk premium for 'New' firms is the largest, irrespective of the values of γ and investment horizons considered. For a \$1 investment in a firm with the highest level of exposure to new technologies, a risk-averse investor with a relative risk aversion of 0.5 would demand an 8.29-cent risk premium to invest their \$1 in the stock for the next year. The same investor would require a 21.59-cent risk premium to invest their \$1 in the stock for the next three years. This pattern suggests that the returns of companies with greater exposure to new technologies exhibit higher levels of uncertainty, prompting risk-averse investors to demand higher risk premia for holding these companies.

There are a few minor points that warrant our attention. Despite the coarse nature of the return distributions, the expected returns for 'New' firms consistently surpass the unconditional expected returns. However, the average return of 'New' firms is dominated by the average return of 'Old' firms for one-year cumulative returns. This discrepancy may be attributed to the omission of other important factors such as firm size as a relevant variable in our analysis.

An additional observation worth noting is the consistent disparity in risk premiums between 'Old' and 'Zero' firms, with the former exhibiting lower values than the latter. This phenomenon may be elucidated by the idea that the complete absence of exposure to novel technologies also carries risks. This paper does not provide explanations to what could be such risks, but one potential reason is the vulnerability of firms to external shocks. Companies that do not invest in innovation may be less resilient in the face of external shocks, such as economic downturns, or regulatory changes as they may lose competitiveness over time. Consequently, a moderate level of engagement with emerging technologies could lead to a less uncertain future.

In summary, the analysis reveals that risk-averse investors demand higher risk premia for companies with greater exposure to new technologies, reflecting higher levels of uncertainty. The finding shows that the performance of the New Minus Old (NMO) portfolio is not solely driven by mispricing, but rather by the uncertainty associated with new technologies. Consequently, the performance of the NMO portfolio reflects the delicate balance between the potential rewards and risks associated with investing in companies with exposure to new technologies.

2.5.2 Exposure to New Technologies and Idiosyncratic Volatility

A fundamental assumption underlying the previous analyses is that exposure to new technologies serves as a proxy for expected future idiosyncratic risks. To validate this premise with data, I calculate idiosyncratic volatility using the residuals from Fama-French three-factor regressions (Fama and French (1993), Fama and French (1996)). Specifically, my analysis concentrates on long-term idiosyncratic volatility, instead of one-month idiosyncratic volatility, due to the extended timeframe often required for the development and uncertainty resolution in new technologies, which typically spans beyond a single month.

At the beginning of each July, the daily excess returns of individual stocks are regressed on the daily three factors, but this is done separately for the first, second, and third subsequent years, each over a one-year period. The idiosyncratic volatility of each stock is then quantified as the standard deviation of the residuals from these annual regressions. Subsequently, these measures are multiplied by 100 to express them as percentages.The three measures of idiosyncratic volatility are matched with the exposure to new technology data available in June of the same year.

Table 9 reports the results. In line with the idea that exposure to emerging technologies correlates with expected idiosyncratic volatility, my findings reveal a positive relationship between exposure to new technologies and future idiosyncratic volatilities. These relationships are statistically significant both cross-sectionally and over time. Furthermore, the largest magnitude and significance of these coefficients are observed in the subsequent one-year idiosyncratic volatility, suggesting that the new technology exposure measure is an effective proxy for predicting expected idiosyncratic volatility over the next year. However, the predictive value of this measure also extends beyond the one-year period, indicating that the uncertainty associated with developing new technologies often spans more than a single year.

2.6 Conclusion

This study examines whether stock returns account for the risks inherent in emerging technologies. New technologies frequently disrupt entire industries and markets; however, the practical viability and commercialization timeline of such technologies often remain uncertain. Using this aspect of emerging technologies, I investigate if the uncertainty in future cash flows, and consequently expected idiosyncratic risks, are priced in a way that is consistent with the asset pricing theories accounting for underdiversified investors.

An extensive dataset comprising all U.S. patent publications and their first- and second-hop neighborhood patents is assembled to detect technology clusters characterized by a high growth rate in new patents. A deep learning model is employed to generate embeddings of patent publications, utilizing both textual data and information on citation networks. These embeddings facilitate the identification of high-growth technology clusters, which are subsequently utilized to define new technology clusters. A firm-level metric of exposure to new technologies is calculated based on patent publication histories over the past three years and the membership of patent publications in relevant technology clusters.

A size-adjusted value-weighted portfolio is constructed by contrasting firms with high exposure to new technology clusters and selling those with low exposure, resulting in a new-minus-old factor (NMO). Between 1981 and 2020, this portfolio generated annual returns of 7.4% and annualized alphas ranging from 5.7% to 14.7%, depending on the factor model employed. In the Fama and MacBeth (1973) regressions of monthly excess returns, the exposure to new technology exhibits a positive loading. Additional analyses are conducted to determine whether the performance of the NMO portfolio could be attributed to risks. The results demonstrate that companies with greater exposure to new technologies display higher levels of uncertainty, leading risk-averse investors to demand higher risk premiums for holding these stocks. Moreover, exposure to new technologies positively relates to future long-term idiosyncratic risks.

Overall, this study contributes to the understanding of the risk-return trade-off investors face when investing in firms engaging in emerging technologies. Also, it contributes to the discussion around idiosyncratic risks and stock prices.

Figure 1. Citation Network and Sample Selection

This figure shows which patents are collected in the citation network of patents. The white nodes are US patent publications and the gray nodes are international patent publications. The circles with a thick outline are collected while the dashed circles are not collected. A gray line between two nodes (circles) shows that the two patents are related because one of the two patent publications cites the other.

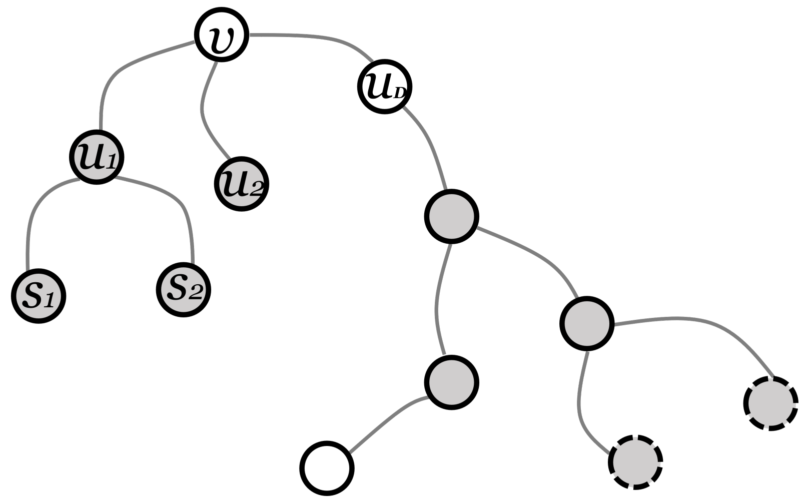


Figure 2. The Deep Learning Model Architecture

This figure shows the deep learning model architecture to compute the embeddings of a U.S. patent publication that will be used for clustering analysis. ❶ Each publication is processed into a sentence that represents it. ❷ A layer of the BERT uncased model translates each token in a sentence into a vector of 768 elements. Therefore, this procedure generates token embeddings of a sentence where the number of tokens varies by sentence. ❸ The token embeddings are averaged out on the same dimension to produce a vector of 768 elements. The vector is then fed into a dense layer with elu activation functions. The resulting vector is used for a skip connection.❹ A layer of an LSTM model takes in a series of a vector of a token in a sentence produced from the BERT layer and outputs a vector of 768 elements that represent the whole sentence. The vector goes through a dense layer with elu activation functions. The resulting vector is concatenated with the vector generated in the third step. ❺ A linear dense layer with elu activation functions that produce a vector of 512 elements. ❻ All vectors of a batch of publications from the previous step and the corresponding network information is fed into a GAT layer with four attention heads. ❼ Skip connection. ❽ A vector from the GAT model augmented by a dense layer with elu activation functions and a vector from the skip connection are concatenated and fed into a dense layer with elu activation functions. ❾ A linear dense layer with elu activation functions that generate a vector of 265 elements for each publication. A dropout rate of 10% is applied to all linear dense layers.

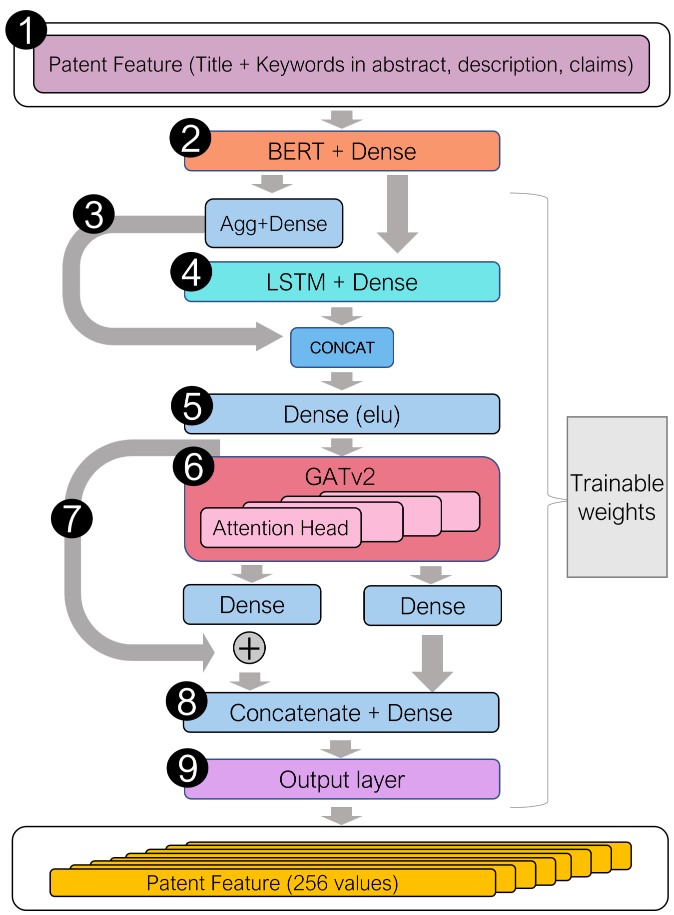


Figure 3. The Exposure to New Technologies of MAMAA Stocks

The figure shows how the exposures to new technologies of MAMAA stocks (Meta Platforms, Amazon, Apple, Microsoft and Alphabet) vary over time. The y-axis represents *NewTechExposure*, a metric quantifying a firm's engagement with emerging technologies on a scale from zero to one. The x-axis displays the corresponding years.

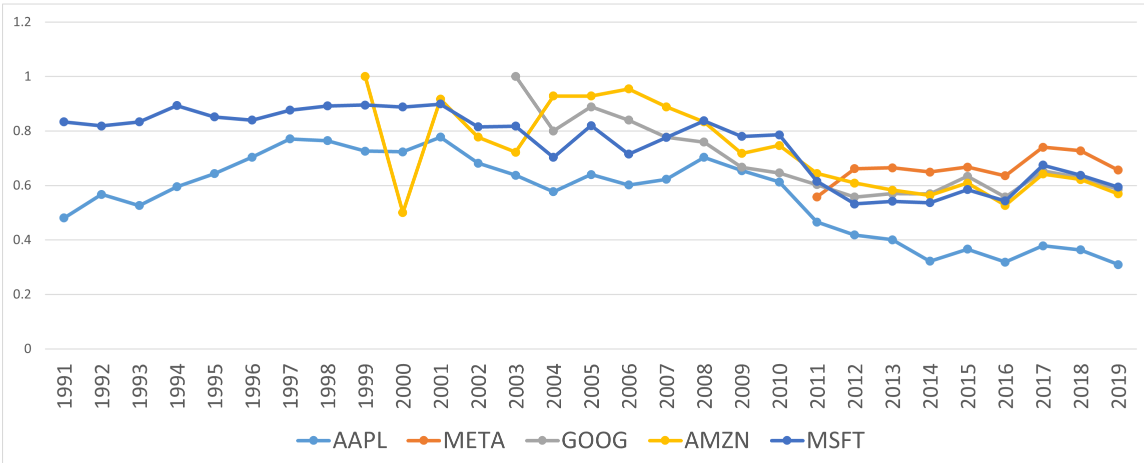


Table 1. Summary Statistics

The table shows the summary statistics of the variables used in the paper. The sample contains U.S. inventions (patent publications publicly available for the first time) between 1981 and 2020 and firms that have at least one invention in the Compustat and CRSP universe. $NewTech$ is an indicator variable that takes one if a patent publication belongs to a new technology cluster at the time it is published, zero otherwise. $cite\_less2y$ is a fraction of patent references published less than two years prior. $cite\_age$ is the median age of patent references. $cite\_age\_std$ is the standard deviation of the age of patent references. $cite\_patentN$ is the log of one plus the number of patent references. $cite\_nonpatentN$ is the log of one plus the number of non-patent references. $cite\_countryN$ is the log of one plus the number of countries of patent references. $cite\_categoryN$ is the log of one plus the number of categories of patent references. $public$ is an indicator variable that takes one if the original assignee of the patent publication is a public firm or a subsidiary of a public firm, zero otherwise. $publicShare$ is a fraction of patent references whose original assignee is a public firm or a subsidiary of a public firm. $cite\_countryHHI$ is a measure of the concentration of countries of patent references, which is calculated by squaring the share of each country and then summing the resulting numbers. $cite\_categoryHHI$ is a measure of the concentration of categories of patent references, which is calculated by squaring the share of a classification of Cooperative Patent Classification and then summing the resulting numbers. $cite\_firmHHI$ is a measure of the concentration of firms of patent references, which is computed by squaring the share of each firm and then summing the resulting numbers. $citedby\_N3y$ is the number of patent publications that cite the invention in three years. $citedby\_N5y$ is the number of patent publications that cite the invention in five years. $NewTechExposure$ is a measure of a firm's exposure to emerging technologies. $InnOrig$ is a measure of the innovative originality of a firm introduced by Hirshleifer, Hsu, and Li (2018). $efficiency$ is a measure of a firm's ability to efficiently generate innovative work defined as the number of patents over R&D capital introduced by Hirshleifer, Hsu, and Li (2013). $R\&Dintensity$ is a measure of R&D intensity measured by R&D expenditures over contemporaneous sales (Chan, Lakonishok, and Sougiannis (2001)). $huge\Delta R\&D$ is a significantly positive abnormal increase in R&D expenditures defined by Eberhart, Maxwell, and Siddique (2004). $InnoAbility$ is a measure of a firm's ability to turn R&D expenditures into future sales introduced by Cohen, Diether, and Malloy (2013). $size$ is the logarithm of total assets. $cash$ is cash and cash equivalent divided by total assets. ml is market leverage. mb is market-to-book ratio. $momentum$ is a measure of a stock's momentum. roa is the return on assets. $forward1y\_ret$ is one-year cumulative future stock returns. $forward3y\_ret$ is three-year cumulative future stock returns. $forward5y\_ret$ is five-year cumulative future stock returns.

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
Invention level variables								
$NewTech$	5754443	0.30	0.46	0.00	0.00	0.00	1.00	1.00
$cite\_less2y$	6039043	0.17	0.19	0.00	0.00	0.11	0.25	1.00
$cite\_age$	6039043	7.72	1.03	0.00	7.30	7.79	8.28	11.09
$cite\_age\_std$	6039043	1.47	0.97	0.00	0.69	1.10	2.26	5.47
$cite\_patentN$	6155936	2.60	0.90	0.00	2.08	2.64	3.09	9.08
$cite\_nonpatentN$	6155936	0.65	1.00	0.00	0.00	0.00	1.10	8.01
$cite\_countryN$	6039043	1.43	0.45	0.69	1.10	1.39	1.79	3.71
$cite\_categoryN$	6031229	1.87	0.66	0.69	1.39	1.79	2.30	6.58
$public$	6155936	0.29	0.45	0.00	0.00	0.00	1.00	1.00
$publicShare$	6039243	0.27	0.24	0.00	0.07	0.22	0.42	1.00
$cite\_countryHHI$	6039043	0.59	0.24	0.08	0.39	0.55	0.76	1.00
$cite\_categoryHHI$	6031229	0.20	0.18	0.00	0.09	0.15	0.25	1.00
$cite\_firmHHI$	6039243	0.42	0.35	0.00	0.15	0.33	0.59	1.00
$citedby\_N3y$	6155936	7.60	20.96	0	1	2	6	200
$citedby\_N5y$	6155936	11.82	30.17	0	2	4	10	300

Table 1-*Continued*

Variable	N	Mean	Std Dev	Minimum	25th Pctl	Median	75th Pctl	Maximum
Firm level variables								
$NewTechExposure$	76477	0.30	0.35	0.00	0.00	0.15	0.50	1.00
$InnOrig$	44916	2.68	0.70	0.00	2.22	2.60	3.07	6.32
$efficiency$	30446	0.21	0.72	0.00	0.00	0.02	0.13	7.89
$R\&Dintensity$	57556	1.16	5.08	0.00	0.02	0.07	0.20	38.27
$huge\Delta R\&D$	57624	0.19	0.39	0.00	0.00	0.00	0.00	1.00
$InnoAbility$	19053	2.40	14.66	-50.15	-1.58	0.65	4.62	78.61
$size$	76477	5.78	2.51	0.35	3.91	5.64	7.54	11.81
$cash$	76477	0.24	0.26	0.00	0.04	0.14	0.37	0.96
ml	72542	0.20	0.22	0.00	0.01	0.12	0.31	0.92
mb	72819	2.47	2.61	0.59	1.16	1.62	2.64	19.17
$momentum$	63134	0.11	0.58	-0.84	-0.23	0.04	0.32	2.74
roa	76477	-0.01	0.36	-1.96	-0.03	0.10	0.16	0.45
$forward1y\_ret$	62907	0.15	0.64	-0.87	-0.23	0.06	0.37	3.22
$forward3y\_ret$	53084	0.44	1.22	-0.96	-0.31	0.19	0.78	6.67
$forward5y\_ret$	43480	0.83	1.84	-0.98	-0.30	0.35	1.28	10.36
$IVol\_1$	86713	3.18	2.27	0.00	1.69	2.58	3.99	82.04
$IVol\_2$	86713	3.19	2.39	0.21	1.67	2.55	3.99	120.28
$IVol\_3$	86713	3.19	2.17	0.14	1.67	2.55	4.02	12.00

Table 2. The Patent Characteristics of New Tech Clusters

The table shows what are the characteristics of patent publications that are classified as relating to new technologies. The sample contains first-time patent publications between 1981 and 2019. The dependent variable ($NewTech$) is an indicator variable that takes one if a patent publication belongs to a new technology cluster at the time it is published, zero otherwise. *cite_less2y* is a fraction of patent references (citations) publicized less than two years before. *cite_age* is the median age of patent references. *cite_age_std* is the standard deviation of the age of patent references. *cite_patentN* is the log of one plus the number of patent references. *cite_nonpatentN* is the log of one plus the number of non-patent references. *cite_countryN* is the log of one plus the number of countries of patent references. *cite_categoryN* is the log of one plus the number of categories of patent references. *public* is an indicator variable that takes one if the original assignee of the patent publication is a public firm or a subsidiary of a public firm, zero otherwise. *publicShare* is a fraction of patent references whose original assignee is a public firm or a subsidiary of a public firm. *cite_countryHHI* is a measure of the concentration of countries of patent references which is calculated by squaring the share of each country and then summing the resulting numbers. *cite_categoryHHI* is a measure of the concentration of categories of patent references which is calculated by squaring the share of a classification of Cooperative Patent Classification and then summing the resulting numbers. *cite_firmHHI* is a measure of the concentration of firms of patent references which is computed by squaring the share of each firm and then summing the resulting numbers. Year and category fixed effects are included in some regression specifications. A category is defined as the first four characters of Cooperative Patent Classification, which consists of a section (one letter A to H and also Y), a class (two digits), and a subclass (one letter). t-statistics based on standard errors clustered at the firm level are shown below the coefficients. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Dep. Var. $NewTech$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
cite_less2y	0.3263***	0.3235***	0.1521***	0.1492***	0.0725***	0.0761***	0.0417***	0.0424***
	[12.7883]	[13.4386]	[19.3439]	[18.3877]	[4.5821]	[4.8954]	[6.0670]	[7.2657]
cite_age					-0.0633***	-0.0620***	-0.0291***	-0.0283***
					[-18.6759]	[-19.4961]	[-16.1970]	[-14.4109]
cite_age_std	0.0073***	0.0068***	0.0042***	0.0039***	0.0101***	0.0095***	0.0061***	0.0057***
	[2.9157]	[2.9114]	[2.7491]	[2.7514]	[4.0420]	[4.0782]	[4.2650]	[4.4168]
cite_patentN	-0.0222**	-0.0256***	-0.0190***	-0.0147***	-0.0333***	-0.0365***	-0.0209***	-0.0172***
	[-2.3654]	[-2.9391]	[-4.9353]	[-3.5785]	[-4.1719]	[-4.9734]	[-5.5023]	[-4.4674]
cite_nonpatentN	0.0337***	0.0353***	0.0161***	0.0199***	0.0308***	0.0324***	0.0154***	0.0192***
	[7.2919]	[7.5973]	[6.3406]	[8.0291]	[7.1526]	[7.4836]	[6.0946]	[7.7974]
cite_countryN	-0.0518***	-0.0485***	-0.0203**	-0.0132	-0.0112	-0.0065	-0.0103	-0.0002
	[-2.7591]	[-2.6650]	[-2.0489]	[-1.5645]	[-0.8777]	[-0.5277]	[-1.4450]	[-0.0426]
cite_categoryN	0.0293**	0.0301***	0.0235***	0.0205***	0.0177	0.0177	0.0230***	0.0193***
	[2.5115]	[2.6727]	[4.3336]	[3.8285]	[1.5156]	[1.5949]	[4.9884]	[4.1852]
public	0.0455***	0.0460***	0.0275***	0.0248***	0.0421***	0.0424***	0.0270***	0.0243***
	[7.4165]	[7.6363]	[6.5238]	[7.0543]	[7.1744]	[7.3873]	[6.5673]	[7.1320]
publicShare	0.2788***	0.2763***	0.0756***	0.0725***	0.2569***	0.2555***	0.0734***	0.0700***
	[6.8698]	[6.9408]	[8.6401]	[9.6164]	[6.9013]	[6.9901]	[8.6430]	[9.8221]
cite_countryHHI					0.0721***	0.0749***	0.0182	0.0239
					[2.5943]	[2.6938]	[1.2059]	[1.4860]
cite_categoryHHI					-0.1033***	-0.1065***	-0.0167	-0.0205**
					[-4.0237]	[-4.1997]	[-1.6028]	[-1.9970]
cite_firmHHI					-0.0474***	-0.0468***	-0.0105***	-0.0091***
					[-4.7699]	[-4.7936]	[-3.5574]	[-4.0615]
Year FE		✓		✓		✓		✓
Category FE			✓	✓			✓	✓
Adjusted R^2	0.073	0.078	0.181	0.188	0.084	0.088	0.182	0.189
Observations	5744749	5744738	5744707	5744696	5744749	5744738	5744707	5744696

Table 3. The Level of the Exposure to New Technologies by Industry

The table presents the number of observations (Obs), mean, and standard deviation (Std. Dev.) of the level of exposure to new technologies for different industries. In addition, the last column reports the proportion of firms in the CRSP database that have information on the level of exposure to new technologies.

Code	Industry	Obs	Mean	Std Dev	Share of Firms
1	Consumer Nondurables – Food, Tobacco, Textiles, Apparel, Leather, Toys	3368	0.15	0.27	27.0
2	Consumer Durables – Cars, TVs, Furniture, Household Appliances	3191	0.17	0.24	54.2
3	Manufacturing – Machinery, Trucks, Planes, Off Furn, Paper, Com Printing	12328	0.14	0.22	52.2
4	Energy Oil, Gas, and Coal Extraction and Products	1720	0.17	0.26	14.3
5	Chemicals and Allied Products	2566	0.14	0.22	54.4
6	Business Equipment – Computers, Software, and Electronic Equipment	18817	0.42	0.36	50.1
7	Telephone and Television Transmission	1385	0.60	0.35	21.6
8	Utilities	969	0.20	0.32	14.6
9	Wholesale, Retail, and Some Services (Laundries, Repair Shops)	3361	0.24	0.34	15.0
10	Healthcare, Medical Equipment, and Drugs	11269	0.33	0.34	56.6
12	Other – Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment	8677	0.33	0.38	20.9

Table 4. The Exposure to New Technologies and Other Innovation Related Variables

The table shows the linear relationship between a firm's exposure to new technologies and other innovation variables investigated by earlier studies. The sample contains firm-year observations between 1981 and 2019. The dependent variable ($NewTechExposure$) is a measure of a firm's exposure to emerging technologies. $InnOrig$ is a measure of the innovative originality of a firm introduced by Hirshleifer, Hsu, and Li (2018). $efficiency$ is a measure of a firm's ability to efficiently generate innovative work defined as the number of patents over R&D capital introduced by Hirshleifer, Hsu, and Li (2013). $R\&Dintensity$ is a measure of R&D intensity measured by R&D expenditure over contemporaneous sales (Chan, Lakonishok, and Sougiannis (2001)). $huge\Delta R\&D$ is the significantly positive abnormal increase in R&D expenditures defined by Eberhart, Maxwell, and Siddique (2004). $InnoAbility$ is a measure of a firm's ability to turn R&D expenditure into future sales introduced by Cohen, Diether, and Malloy (2013). $size$ is the logarithm of total assets. $cash$ is cash and cash equivalent divided by total assets. ml is market leverage. mb is market-to-book ratio. In all regressions, industry (defined as the first three digits of SIC) times year fixed effects are included. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Dep. Var. NewTechExposure									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$InnOrig$	0.0079**					0.0033				
	[2.1506]					[0.8838]				
$efficiency$		-0.0092***					-0.0084***			
		[-2.9880]					[-2.9616]			
$R\&Dintensity$			0.0008*					-0.0008*		
			[1.8468]					[-1.8655]		
$huge\Delta R\&D$				0.0167***					0.0145***	
				[4.6137]					[4.0470]	
$InnoAbility$					-0.0003**					-0.0003*
					[-2.0468]					[-1.8086]
$size$						0.0047***	0.0036**	0.0023*	0.0023**	0.0073***
						[3.6551]	[2.4330]	[1.9138]	[1.9633]	[3.7882]
$cash$						0.1311***	0.1137***	0.1386***	0.1251***	0.1106***
						[10.6458]	[7.5391]	[11.7779]	[11.1983]	[4.6461]
ml						-0.0225**	-0.0337**	-0.0337***	-0.0296**	-0.0286
						[-2.0064]	[-2.4279]	[-2.8316]	[-2.5088]	[-1.2769]
mb						0.0035***	0.0048***	0.0023**	0.0023***	0.0018
						[3.3564]	[3.7074]	[2.3836]	[2.5790]	[0.8679]
Adjusted R^2	0.433	0.437	0.354	0.361	0.308	0.439	0.443	0.363	0.369	0.314
Observations	42686	28781	55469	55789	17717	41598	28361	52784	52934	17486

Table 5. Monthly Returns of New Tech Exposure Portfolios

The table shows the monthly returns of new tech exposure portfolios. Each year, portfolios are formed based on firms' exposure to new technologies at the end of June of year t from 1981 to 2019. The portfolio 'No' is formed with firms whose exposure to new technologies is zero. With firms whose exposure to new technologies is positive, three portfolios ('Low', 'Middle', 'High') are formed based on the 30th and 70th percentiles of the exposure. Firms are further sorted independently based on the NYSE median size breakpoint at the end of June of year t. I compute monthly size-adjusted returns of exposure portfolios ('Zero', 'Old', 'Mid', 'New') by computing the average of double-sorted portfolios. For example, the portfolio 'New' is (High/Small+High/Big)/2. All portfolios are value-weighted portfolios where weights are market capitalization at the end of June of year t. t-statistics are shown in the square brackets. All returns are expressed in decimals.

New Tech Exposure Portfolio	All	Small	Big	Size-Adjusted New Tech Exposure Portfolio	
No (excess returns)	0.0133	0.0246	0.0118	Zero (excess returns)	0.0182
	[6.31]	[8.42]	[5.66]		[7.69]
Low (excess returns)	0.0113	0.0263	0.0109	Old (excess returns)	0.0186
	[5.78]	[7.60]	[5.61]		[7.47]
Middle (excess returns)	0.0134	0.0299	0.0129	Mid (excess returns)	0.0214
	[5.37]	[7.41]	[5.19]		[7.00]
High (excess returns)	0.0171	0.0337	0.0154	New (excess returns)	0.0246
	[6.50]	[6.89]	[6.08]		[7.08]
High-No	0.0037	0.0091	0.0037	New minus Zero (NMZ)	0.0064
	[2.30]	[2.74]	[2.27]		[2.96]
High-Low	0.0058	0.0075	0.0045	New minus Old (NMO)	0.0060
	[3.43]	[2.39]	[2.81]		[2.99]

Table 6. Alphas of New Tech Exposure Portfolios

The table shows the alphas, t-statistics, and R^2 of new tech exposure portfolios from time-series regressions of monthly returns with different factors. Eight different models are considered: Four-factor (market, size, value, momentum factors) model (4F). 4F plus the robust-minus-weak (RMW) factor and the conservative-minus-aggressive (CMA) factor model. The following portfolios are augmented to the 4F plus RMW+CMA model. ORIG: monthly returns of a portfolio that contrasts firms with high originality and firms with low originality, as introduced by Hirshleifer, Hsu, and Li (2018). EFF: A portfolio that takes long positions in firms with high innovation efficiency and short positions in firms with low efficiency, as constructed by Hirshleifer, Hsu, and Li (2013). INT: A portfolio adopting long positions in firms with high R&D intensity and short positions in firms with low R&D intensity measured by (Chan, Lakonishok, and Sougiannis (2001)). INC: a strategy taking long positions in firms that experience huge increases in R&D expenditure identified by Eberhart, Maxwell, and Siddique (2004) and short positions in the 3-month treasury bill. ABI: long firms with high R&D expenditure and high ability to turn it into sales and short firms with high R&D expenditure and low ability measured by Cohen, Diether, and Malloy (2013). In the last column, all factors are included. All alphas are expressed in decimals. t-statistics are shown in the square brackets.

Alphas from different factor models								
		4F plus	4F + RMW+CMA plus					
	4F	RMW+CMA	ORIG	EFF	INT	INC	ABI	ALL 5
New (excess returns)	0.0181	0.0207	0.0207	0.0221	0.0133	0.0156	0.0211	0.0150
	[10.37]	[12.14]	[12.28]	[13.91]	[7.91]	[8.38]	[11.91]	[9.08]
Zero (excess returns)	0.0111	0.0100	0.0100	0.0091	0.0103	0.0090	0.0104	0.0084
	[16.93]	[15.52]	[15.65]	[12.21]	[14.50]	[12.43]	[15.59]	[9.69]
Old (excess returns)	0.0118	0.0112	0.0112	0.0106	0.0087	0.0073	0.0110	0.0050
	[13.50]	[12.75]	[12.75]	[9.47]	[9.38]	[8.07]	[13.09]	[4.65]
New minus Zero (NMZ)	0.0070	0.0107	0.0107	0.0130	0.0030	0.0066	0.0107	0.0066
	[3.69]	[6.00]	[6.10]	[7.88]	[1.72]	[3.34]	[5.80]	[3.85]
New minus Old (NMO)	0.0063	0.0096	0.0095	0.0115	0.0047	0.0082	0.0101	0.0100
	[3.41]	[5.29]	[5.35]	[6.85]	[2.45]	[4.03]	[5.46]	[5.34]

R^2 of different factor models								
		4F plus	4F + RMW+CMA plus					
	4F	RMW+CMA	ORIG	EFF	INT	INC	ABI	ALL 5
New (excess returns)	0.77	0.79	0.80	0.88	0.83	0.81	0.79	0.91
Zero (excess returns)	0.93	0.94	0.94	0.93	0.94	0.94	0.93	0.94
Old (excess returns)	0.88	0.89	0.89	0.88	0.90	0.91	0.90	0.92
New minus Zero (NMZ)	0.28	0.40	0.43	0.63	0.52	0.43	0.41	0.74
New minus Old (NMO)	0.22	0.30	0.33	0.51	0.36	0.31	0.31	0.58

Table 7. Return Predictive Power of New Tech Exposure

The table shows the predictive power of the exposure to new technologies from monthly Fama and MacBeth (1973) cross-sectional regressions of monthly excess returns. The sample period is from July 1982 to June 2020. $NewTechExposure$ is a measure of a firm's exposure to emerging technologies. $MktCap$ is the logarithm of market capitalization. mb is market-to-book ratio. Operating profitability (op) is revenues minus cost of goods sold, minus selling, general, and administrative expenses, minus interest expense all divided by book equity. Investment (inv) is the change in total assets from the fiscal year ending in year $t-2$ to the fiscal year ending in $t-1$, divided by $t-2$ total assets. $momentum$ is a measure of a stock's momentum measured at the end of June based on returns over the previous 6 months. $R\&Dintensity$ is a measure of R&D intensity measured by R&D expenditure over contemporaneous sales (Chan, Lakonishok, and Sougiannis (2001)). $huge\Delta R\&D$ is the significantly positive abnormal increase in R&D expenditures defined by Eberhart, Maxwell, and Siddique (2004). The estimated coefficients are reported in percentages. R^2 and observations are the time-series average from the monthly cross-sectional regressions. t-statistics based on Newey and West (1986) autocorrelation-adjusted heteroscedasticity-robust standard errors are shown in square brackets.

	Dep. Var. Excess Returns						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$NewTechExposure$	0.376	0.420	0.331	0.441	0.472	0.400	0.431
	[1.87]	[2.07]	[1.72]	[2.15]	[2.28]	[2.02]	[2.17]
$MktCap$	0.332	0.351	0.338	0.389	0.360	0.397	0.363
	[7.47]	[8.7]	[7.65]	[8.94]	[9.05]	[9.18]	[9.07]
mb	-0.102	-0.135	-0.105	-0.150	-0.162	-0.153	-0.163
	[-3.74]	[-4.93]	[-3.89]	[-6.22]	[-6.26]	[-6.36]	[-6.20]
op		-0.234			-0.212		-0.217
		[-3.70]			[-3.28]		[-3.37]
inv		-0.331			-0.355		-0.373
		[-5.42]			[-5.15]		[-5.43]
$momentum$			-0.514			-0.592	-0.468
			[-3.47]			[-4.04]	[-3.09]
$huge\Delta R\&D$				0.329	0.210	0.331	0.204
				[3.08]	[1.92]	[3.16]	[1.92]
$R\&Dintensity$				0.060	0.157	0.067	0.160
				[1.91]	[1.64]	[2.1]	[1.65]
R^2	0.03	0.03	0.03	0.03	0.04	0.04	0.05
$Observations$	1521	1228	1514	1123	917	1118	913

Table 8. Risk Premium from Power Utility Function

The table shows the risk premiums extracted from power utility functions in the context of investing in stocks of firms with different levels of exposure to new technologies. The risk premium (RP) is defined by the following equation: $E[u(W)] = u(E[W] - RP)$. For a constant γ with $0 < \gamma < 1$, u(W) = $\frac{1}{1-\gamma}W^{1-\gamma}$. W is defined as a dollar amount of wealth when a dollar is invested in a stock of a firm with a certain level of exposure to new technologies. Firms under the heading of 'Zero' are the firms with zero value of $NewTechExposure$. With firms whose exposure to new technologies is positive, three groups ('Old', 'Mid', 'New') are formed based on the 30th and 70th percentiles of $NewTechExposure$. Return distributions are provided in Appendix E.

	One-year cumulative returns					
	γ	Unconditional	Zero	Old	Mid	New
Expected return		1.1193	1.1020	1.1324	1.1269	1.1286
	0.3	0.0413	0.0416	0.0313	0.0433	0.0499
Risk Premium	0.5	0.0685	0.0690	0.0522	0.0721	0.0829
	0.7	0.0959	0.0965	0.0732	0.1003	0.1160
	Three-year cumulative returns					
	γ	Unconditional	Zero	Old	Mid	New
Expected return		1.4089	1.3923	1.4018	1.4243	1.4096
	0.3	0.1122	0.1134	0.0836	0.1202	0.1310
Risk Premium	0.5	0.1838	0.1858	0.1383	0.1963	0.2159
	0.7	0.2557	0.2583	0.1940	0.2721	0.2992

Table 9. New Tech Exposure and Future Idiosyncratic Volatility

The table shows the relationship between exposure to new technologies and subsequent idiosyncratic volatility. The sample period is from 1981 to 2019. $NewTechExposure$ is a measure of a firm's exposure to emerging technologies. $Ivol\_K$, where $K \in \{1, 2, 3\}$, represents the standard deviation of residuals from the Fama-French three-factor model. In this model, daily excess returns are regressed on market, size, and value factors over the subsequent Kth year. All idiosyncratic volatilities are scaled up by a factor of 100 to convert them into percentage terms. In the last three regressions, control variables are included but their coefficients are not reported. The control variables are firm size, market-to-book ratio, market leverage, and the level of cash holdings. t-statistics based on standard errors clustered at the firm level are shown below the coefficients. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	IVol_1	IVol_2	IVol_3	IVol_1	IVol_2	IVol_3
$NewTechExposure$	0.670***	0.614***	0.561***	0.228***	0.197***	0.202***
	[12.35]	[10.81]	[11.05]	[5.00]	[4.42]	[5.56]
Year FE	✓	✓	✓			
Firm FE				✓	✓	✓
Adjusted R^2	0.117	0.105	0.124	0.514	0.513	0.598
Observations	48779	48779	48779	48157	48157	48157

	IVol_1	IVol_2	IVol_3	IVol_1	IVol_2	IVol_3
$NewTechExposure$	0.193***	0.126***	0.100***	0.202***	0.129***	0.098***
	[4.65]	[3.06]	[3.10]	[5.11]	[3.20]	[3.09]
Year FE	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Controls				✓	✓	✓
Adjusted R^2	0.596	0.586	0.683	0.614	0.591	0.688
Observations	48157	48157	48157	47859	47859	47859

Online Appendix

New Technologies and Stock Returns

Jinyoung Kim

Appendix A. Matching assignees to firms

In this section, I explain the procedure that matches assignees to firms. I collect the first (original) assignees' information from Google Patent. I gather CIK and historical company names registered with the SEC from the SEC bulk data set. Subsidiary names are obtained by cleaning the subsidiary data provided by WRDS between 1995 and 2019. To gather subsidiary data prior to 1995 and after 2019, I collect Exhibit 21 of Annual Report 10-K. Whenever there is no subsidiary information, I infer that the subsidiaries of a firm are the same as those in the previous year.

To match assignees of a patent or application to US companies, I eliminate company abbreviations, regional attributes, and legal marks from assignee information as well as company information, and find the exact match. The matching outcomes are presented in Table B. I compare my results with those of Kogan et al. (2017). One important difference between their matching method and mine is that they use a fuzzy name-matching method while I conduct the exact name-matching. Therefore, I have fewer patents matched to CRSP firms. However, I not only use parent firms' names but also use their subsidiary information, which allows me to have more CRSP firms found in the patent data.

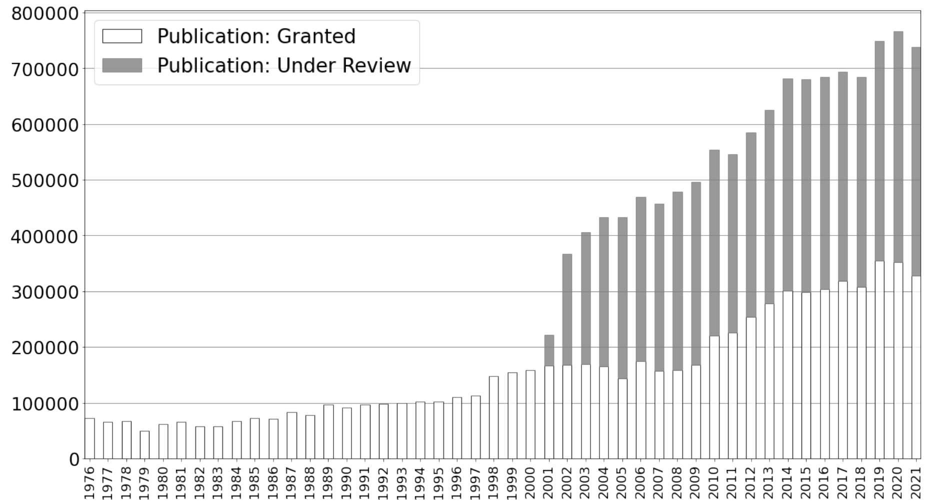
Table A. Matching Assignees to Firms

pubyear	Publications	Publications - Granted	Publications - Review	Kogan et al. (2017) Table A.2
1970_1979	663,335	663,335	-	690,459
1980_1989	713,084	713,084	-	708,735
1990_1999	1,115,140	1,115,140	-	1,109,398
2000_2010	4,472,929	1,849,919	2,623,010	1,846,063
2011_2021	7,432,689	3,323,691	4,108,998	-
Total	14,397,177	7,665,169	6,732,008	4,354,655
pubyear	Matched Publications	Matched Pub. - Granted	Matched Pub. - Review	Kogan et al. (2017) Table A.2
1970_1979	161,096	161,096	-	247,102
1980_1989	162,693	162,693	-	235,525
1990_1999	305,935	305,935	-	352,005
2000_2010	1,457,708	646,843	810,865	729,324
2011_2021	2,271,135	1,098,722	1,172,413	-
Total	4,358,567	2,375,289	1,983,278	1,563,956
pubyear	Matched PERMCO	Matched PERMCO -Granted	Matched PERMCO -Reviewed	Kogan et al. (2017) Table A.2
1970_1979	2,390	2,390	-	2,086
1980_1989	3,059	3,059	-	2,756
1990_1999	4,903	4,903	-	3,664
2000_2010	6,723	5,631	5,854	4,415
2011_2021	5,906	5,319	5,506	-
Total	10,655	10,005	7,486	-

Appendix B. Figures Related to the Patent Publications

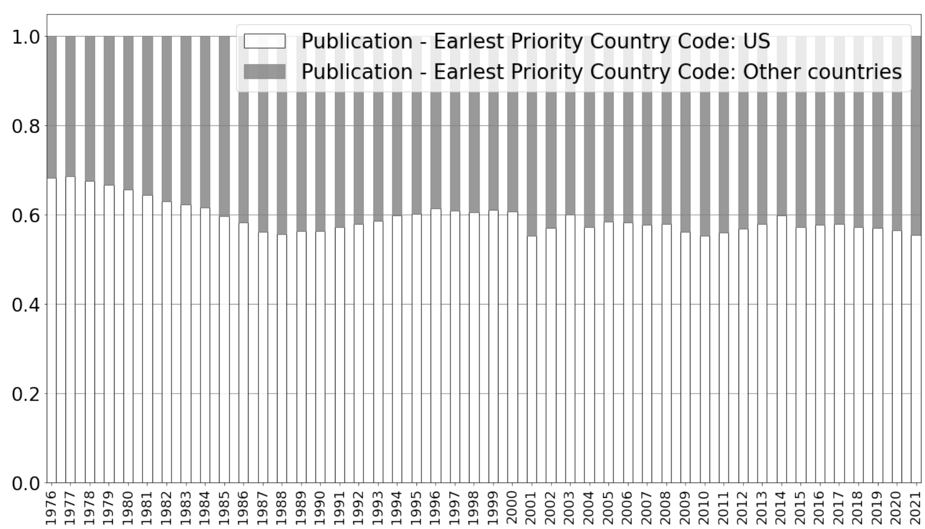
Fiqure B1. The Number of Publications

This figure shows the number of patent publications in the U.S. The gray bar is the number of publications that are under review, and the white bar is the number of publications about patents granted. The x-axis is the year in which patent publications are publicized.



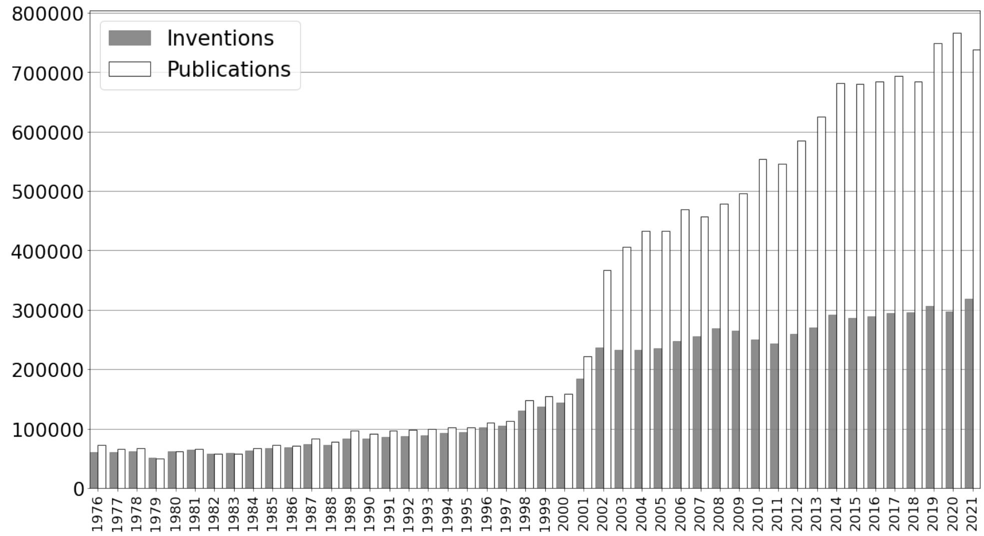
Fiqure B2. The Origin of Publications

This figure shows the origin of U.S. patent publications. The white bar is the share of publications whose country code of the earliest priority code is the US. The gray bar is the share of publications whose country code of the earliest priority code is the other country code.



Fiqure B3. The Number of Inventions vs. Publications - U.S.

This figure shows the number of inventions and patent publications in the U.S. Inventions are identified as the earliest priority code. If there's no priority code, inventions are the earliest publication code. Publications are sorted by publication year, and inventions are sorted by the year of the first publication.



Fiqure B4. The Number of Inventions vs. Publications - First-hop neighbors

This figure shows the number of unique inventions and unique patent publications of patents that are in the first-hop neighborhood of US patent publications. Inventions are identified as the earliest priority code. If there's no priority code, inventions are the earliest publication code. Publications are sorted by publication year, and inventions are sorted by the year of the first publication.

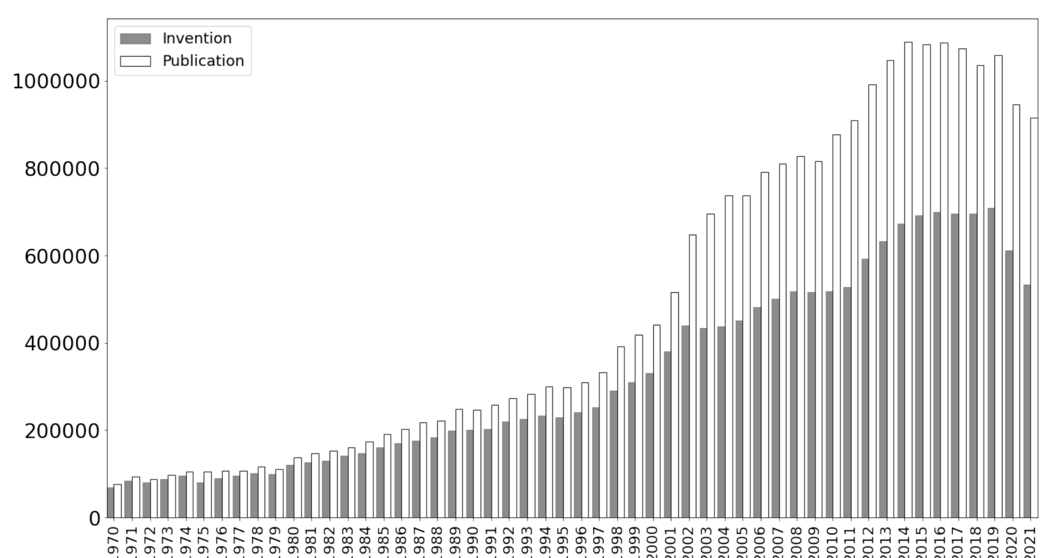


Figure B5. The Number of Publications that cite U.S. publications

This figure shows the number of unique publications that cite U.S. publications sorted by the year in which they are published. The publications are grouped by the country code of the earliest prior code.

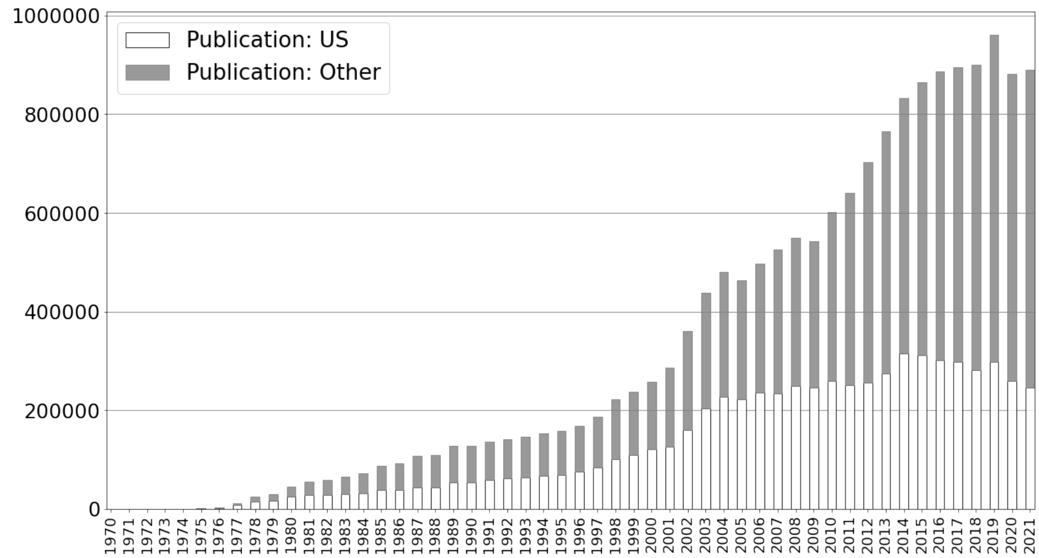
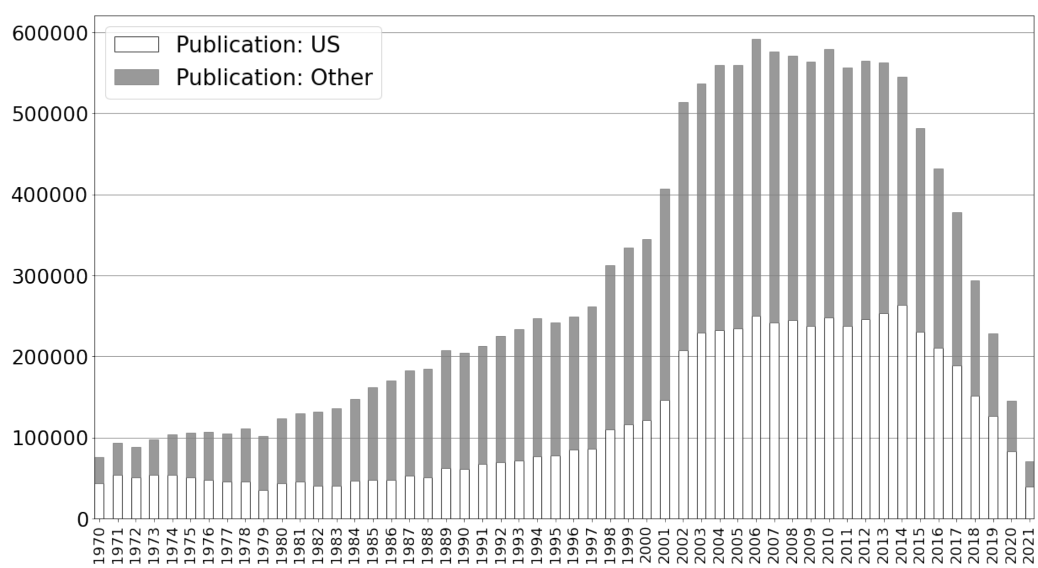


Figure B6. The Number of Publications that are cited by U.S. publications

This figure shows the number of unique publications that are cited by U.S. publications sorted by the year in which they are published. The publications are grouped by the country code of the earliest prior code.



Appendix C. Stock Returns and Scientific Contribution of New Tech Firms in the Long Run

The performance of the new minus old factor and the positive loading of the new tech exposure may indicate the risky nature of emerging technologies. However, as pointed out by previous studies (Chan, Lakonishok, and Sougiannis (2001), Eberhart, Maxwell, and Siddique (2004), Hirshleifer, Hsu, and Li (2013), Hirshleifer, Hsu, and Li (2013), Cohen, Diether, and Malloy (2013), Leung, Evans, and Mazouz (2020), Fitzgerald et al. (2021)), firms' innovative activities are highly difficult to process. Therefore, there is always a component relating to mispricing when it comes to factors generated from firms' technological investments. However, that does not mean that all of the results in the paper are driven by mispricing. In this section, I investigate the nature of the investment in emerging technologies by examining companies' stock performance in the long run as well as invention-level performances.

C.1 Long-Term Stock Returns of New Tech Firms

In this subsection, I investigate the stock returns of firms sorted by the level of exposure to new technologies. The idea behind the investigation of the long-term performance of firms grouped by the level of exposure to emerging technologies is that uncertainty associated with new technologies is likely to be resolved in the long run and if the finding in the previous section is only driven by mispricing, we should observe that the firms with high exposure to new technologies outperform other firms in the future to some extent.

The classification of firms into 'No', 'Low', 'Middle', and 'High' group of new tech exposure is the same as in the previous section. Firm-year observations where the variable $NewTechExposure$ is zero are in the 'No' exposure group. Those below the 30th percentile of the variable are in the 'Low' exposure group. Those above the 70th percentile belong to the 'High' exposure group. The rest are the 'Middle' group. At the beginning of each year, I compute three measures of long-term stock returns using the information of $NewTechExposure$ available at the time. Forward K-year cumulative returns are computed as follows.

$$\prod_{m=1}^{K\times 12}(R_{i,\tau+m}+1)-1 \tag{8}$$

where τ is a month (December) when the information about the exposure to new technologies is measured. $R_{i,\tau+m}$ is the monthly return of stock i in month $\tau + m$ in decimal form. All long-term stock returns are Winsorized at 1% and 99% levels.

The statistics of the stock returns are presented in Table C1. First, the mean value of long-term stock returns does not exhibit a notable pattern when sorted by the level of exposure to emerging technologies. Second, the values of standard deviation, however, show a clear pattern. Across all windows of stock returns, the standard deviation of the cumulative returns is always larger for firm-year observations in the 'High' exposure group than for observations in the 'Low' exposure group. When comparing the 'High' group and the 'No' group, the standard deviation is always larger for the 'High' group. The findings indicate that firms with high exposure to new technologies are experiencing a more uncertain future.

Further investigation into the percentiles supports the findings. The value of the percentiles of all types of cumulative returns up to median show that the numbers are always smaller for firms in the 'High' group. However, starting 75th percentile, firms in the 'High' group show better (or at least second-to-none) long-term stock performance than other groups. The findings show that even though firms with high exposure to new technologies have higher chances of performing worse than other firms, once it succeed, the reward is going to be higher than what the rest get.

The findings in this subsection highlight that mispricing is not the sole driver of the staggering performance of new-minus-old factors or the positive loading of new tech exposure in the monthly stock return regressions. It is the risk-return trade-off that contributes to the findings in the previous section. If it is to be the outcome of mispricing, we should observe the persistent, superior long-term performance of firms with high exposure to new technologies as with Eberhart, Maxwell, and Siddique (2004).

C.2 Forward Citations of Inventions

In this subsection, I investigate forward citations of inventions and compare inventions classified as new technologies with those not. Forward citations are citations that a focal patent (an invention) receives over time. When conducting analysis with forward citations of patents, we are faced with truncation bias problems (see Hall, Jaffe, and Trajtenberg (2001), Dass, Nanda, and Xiao (2017), and Lerner and Seru (2022)). In order to overcome the problem, I compute the number of forward citations by setting a specific window over which the number of citations is counted. I look at a three-year window and a five-year window. Furthermore, I restrict the sample period accordingly so all inventions have full three- or five-year worth of citations.

Table C2 reports the summary statistics of forward citations in the sample period. Unlike long-term stock returns, there is a stark difference in the mean values of the number of forward citations of new tech inventions and non-new tech inventions. The number is always larger for inventions relating to new technologies. The percentiles also deliver similar information; inventions in new technologies have more scientific contributions in the future. However, similar to what is found in the previous subsection, the standard deviation of the number of forward citations is always greater for new tech inventions. The finding implies that it is more difficult to tell which invention is going to be more influential if two inventions are new technologies.

The finding that inventions relating to emerging technologies are cited more going forward adds to the literature that searches for the attribute of innovative work that produces more scientific contributions. For example, the degree of distance-to-science (Fleming and Sorenson (2004), Ahmadpoor and Jones (2017), Watzinger and Schnitzer (2019)) or the similarity to other patents (Kelly et al. (2021)) may determine the level of scientific contribution of an invention. I further show that when a patent belongs to a new technological group, it is more likely to be influential in the technological landscape.

The level of scientific contribution is measured by the number of citations an invention received in a given time window.

The positive coefficients indicate that inventions that are related to growing technologies are more likely to receive citations. This finding is consistent with the finding of Acemoglu, Akcigit, and Kerr (2016) that the patent growth in upstream technology has strong predictive power on future downstream innovation.

Table C1. Future Stock Returns by the Level of New Tech Exposure

The table shows the summary statistics of future stock returns of firms sorted by the level of exposure to new technologies at a given point in time. The information on the exposure is measured at the beginning of the year when I start to compute cumulative returns. For example, forward three-year cumulative returns are computed as follows. $\prod_{m=1}^{36}(R_{i,\tau+m} + 1) - 1$ where τ is a month (December) when the information about the exposure to new technologies is measured. $R_{i,\tau+m}$ represents the monthly return of stock i in month $\tau + m$ in decimal form. All variables are Winsorized at 1% and 99% levels.

Analysis Variable: forward 3-year cumulative returns								
Exposure	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
No	19969	0.427	1.223	-0.960	-0.335	0.164	0.776	6.500
Low	10823	0.451	1.044	-0.929	-0.163	0.263	0.773	4.983
Middle	13279	0.451	1.278	-0.960	-0.332	0.171	0.773	6.669
High	9013	0.446	1.310	-0.960	-0.383	0.135	0.807	6.669
Analysis Variable: forward 5-year cumulative returns								
Exposure	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
No	16424	0.824	1.845	-0.977	-0.308	0.333	1.260	10.181
Low	9129	0.845	1.629	-0.959	-0.123	0.468	1.287	8.337
Middle	10813	0.829	1.916	-0.977	-0.360	0.319	1.282	10.360
High	7114	0.801	1.973	-0.977	-0.449	0.237	1.287	10.360

Table C2. Forward Citations of Inventions: New Tech vs. Those Not

The table shows the summary statistics of forward citations of inventions identified as relating to new technologies at a given point in time. $NewTech$ is an indicator variable that takes one if an invention belongs to a new technology cluster at the time it is published, and zero otherwise. In order to avoid truncation bias, I limit the sample periods accordingly.

Analysis Variable: Forward citations in three years, 1981-2017								
$NewTech$	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
0	4168435	5.772	16.669	0	1	2	5	78
1	1653764	9.057	22.660	0	1	3	7	136
Analysis Variable: Forward citations in three years, 2001-2017								
$NewTech$	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
0	2757259	7.132	19.422	0	1	2	6	104
1	1263543	10.121	24.620	0	1	3	8	160
Analysis Variable: Forward citations in five years, 1981-2015								
$NewTech$	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
0	3611803	8.876	22.879	0	1	4	8	103
1	1433569	13.918	31.768	0	2	5	12	178
Analysis Variable: Forward citations in five years, 2001-2015								
$NewTech$	N	Mean	Std Dev	1st Pctl	25th Pctl	Median	75th Pctl	99th Pctl
0	2200627	11.004	27.254	0	2	4	10	137
1	1043348	15.385	34.665	0	2	6	14	209

Appendix D. Forward Citations - Regressions

Table D1. Three-Year Forward Citations and Patent Characteristics

The table shows the relationship between forward citations and patent characteristics. The dependent variable is the log of one plus the number of patents that cite the focal invention in three years after the invention was first published. $NewTech$ is an indicator variable that takes one if a patent publication belongs to a new technology cluster at the time it is published, zero otherwise. $cite\_less2y$ is a fraction of patent references (citations) publicized less than two years before. $cite\_age$ is the median age of patent references. $cite\_age\_std$ is the standard deviation of the age of patent references. $cite\_patentN$ is the log of one plus the number of patent references. $cite\_nonpatentN$ is the log of one plus the number of non-patent references. $cite\_countryN$ is the log of one plus the number of countries of patent references. $cite\_categoryN$ is the log of one plus the number of categories of patent references. $publicShare$ is a fraction of patent references whose original assignee is a public firm or a subsidiary of a public firm. $cite\_countryHHI$ is a measure of the concentration of countries of patent references. $cite\_categoryHHI$ is a measure of the concentration of categories of patent references. $cite\_firmHHI$ is a measure of the concentration of firms of patent references. Year and category fixed effects are included in some regression specifications. Category is defined as the first four characters of Cooperative Patent Classification. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Dep. Var. Log of the Number of Forward Citations in 3 Years							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewTech$	0.2957***	0.2504***	0.1290***	0.1262***	0.1611***	0.1642***	0.0945***	0.0971***
	[9.8343]	[10.5324]	[11.0440]	[11.9671]	[10.8879]	[11.7096]	[13.5853]	[13.7012]
$cite\_less2y$					-0.2477***	-0.2528***	-0.2751***	-0.2814***
					[-12.5596]	[-12.9601]	[-16.6317]	[-18.0052]
$cite\_age$					-0.1224***	-0.1190***	-0.0777***	-0.0761***
					[-12.2194]	[-12.1386]	[-9.5445]	[-9.3880]
$cite\_age\_std$					0.0692***	0.0656***	0.0705***	0.0672***
					[13.2730]	[12.8365]	[15.2224]	[15.1117]
$cite\_patentN$					0.3439***	0.3336***	0.3202***	0.3153***
					[17.3083]	[17.0045]	[26.4348]	[26.0236]
$cite\_nonpatentN$					0.1295***	0.1306***	0.1212***	0.1251***
					[21.0172]	[21.0366]	[24.4635]	[24.3548]
$cite\_countryN$					0.2317***	0.2357***	0.2522***	0.2614***
					[5.7902]	[5.8438]	[10.2350]	[10.6450]
$cite\_categoryN$					0.0469**	0.0476**	0.0966***	0.0936***
					[2.4013]	[2.4589]	[9.5273]	[9.4635]
$publicShare$					0.3947***	0.3722***	0.2170***	0.1982***
					[16.4780]	[16.0597]	[13.9700]	[12.8498]
$cite\_countryHHI$					0.3437***	0.3387***	0.3756***	0.3743***
					[5.8938]	[5.7839]	[11.7913]	[11.7340]
$cite\_categoryHHI$					0.1632***	0.1557***	0.3058***	0.2935***
					[4.8774]	[4.6069]	[20.0420]	[19.3873]
$cite\_firmHHI$					-0.0754***	-0.0717***	-0.0426***	-0.0395***
					[-11.4217]	[-11.0561]	[-8.3774]	[-7.9045]
Year FE		✓		✓		✓		✓
Firm FE			✓	✓			✓	✓
Adjusted R^2	0.015	0.054	0.065	0.089	0.182	0.188	0.203	0.209
Observations	6377498	6377498	6377453	6377453	5642664	5642664	5642621	5642621

Table D2. Five-Year Forward Citations and Patent Characteristics

The table shows the relationship between forward citations and patent characteristics. The dependent variable is the log of one plus the number of publications that cite the focal invention in five years after the invention was first published. $NewTech$ is an indicator variable that takes one if a patent publication belongs to a new technology cluster at the time it is published, zero otherwise. $cite\_less2y$ is a fraction of patent references (citations) publicized less than two years before. $cite\_age$ is the median age of patent references. $cite\_age\_std$ is the standard deviation of the age of patent references. $cite\_patentN$ is the log of one plus the number of patent references. $cite\_nonpatentN$ is the log of one plus the number of non-patent references. $cite\_countryN$ is the log of one plus the number of countries of patent references. $cite\_categoryN$ is the log of one plus the number of categories of patent references. $publicShare$ is a fraction of patent references whose original assignee is a public firm or a subsidiary of a public firm. $cite\_countryHHI$ is a measure of the concentration of countries of patent references. $cite\_categoryHHI$ is a measure of the concentration of categories of patent references. $cite\_firmHHI$ is a measure of the concentration of firms of patent references. Year and category fixed effects are included in some regression specifications. Category is defined as the first four characters of Cooperative Patent Classification. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Dep. Var. Log of the Number of Forward Citations in 5 Years							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$NewTech$	0.3104***	0.2617***	0.1496***	0.1362***	0.1655***	0.1659***	0.1110***	0.1025***
	[10.3800]	[10.4970]	[12.5584]	[11.5706]	[12.2405]	[12.1410]	[12.5753]	[13.1875]
$cite\_less2y$					-0.1812***	-0.1801***	-0.1999***	-0.2024***
					[-7.8638]	[-7.6612]	[-10.5967]	[-11.7041]
$cite\_age$					-0.1312***	-0.1301***	-0.0920***	-0.0901***
					[-11.7764]	[-11.7516]	[-10.1529]	[-9.7239]
$cite\_age\_std$					0.0815***	0.0766***	0.0831***	0.0786***
					[14.5995]	[14.1872]	[16.5434]	[16.8846]
$cite\_patentN$					0.3670***	0.3644***	0.3390***	0.3435***
					[19.1915]	[18.8645]	[28.1275]	[28.1052]
$cite\_nonpatentN$					0.1270***	0.1337***	0.1260***	0.1349***
					[18.2152]	[18.1569]	[22.9196]	[24.2978]
$cite\_countryN$					0.2164***	0.2429***	0.2247***	0.2566***
					[5.0171]	[5.7746]	[8.4443]	[10.4491]
$cite\_categoryN$					0.0368**	0.0302	0.0891***	0.0776***
					[1.9779]	[1.5857]	[9.1002]	[7.9822]
$publicShare$					0.3775***	0.3539***	0.2283***	0.1927***
					[17.2601]	[16.2257]	[17.3256]	[13.6424]
$cite\_countryHHI$					0.3304***	0.3356***	0.3443***	0.3519***
					[5.3500]	[5.5299]	[10.7753]	[11.6009]
$cite\_categoryHHI$					0.1260***	0.0997***	0.2718***	0.2476***
					[3.4500]	[2.7952]	[16.3243]	[15.8931]
$cite\_firmHHI$					-0.0698***	-0.0651***	-0.0417***	-0.0350***
					[-10.1494]	[-9.9299]	[-8.1846]	[-6.9810]
Year FE		Y		Y		Y		Y
Firm FE			Y	Y			Y	Y
Adjusted R^2	0.016	0.055	0.064	0.091	0.183	0.198	0.202	0.218
Observations	6377498	6377498	6377453	6377453	5642664	5642664	5642621	5642621

Appendix E. Return Distribution

Table E1. One-Year Return Distribution

One-year return distribution													
		All	Zero	Old	Mid	New			All	Zero	Old	Mid	New
R	W	P	P	P	P	P	R	W	P	P	P	P	P
-0.808	0.192	0.016	0.016	0.009	0.015	0.026	0.954	1.954	0.004	0.003	0.005	0.004	0.003
-0.790	0.210	0.006	0.006	0.003	0.007	0.007	0.989	1.989	0.004	0.004	0.003	0.004	0.005
-0.755	0.245	0.007	0.008	0.005	0.007	0.009	1.025	2.025	0.003	0.003	0.002	0.003	0.004
-0.719	0.281	0.008	0.007	0.005	0.008	0.011	1.060	2.060	0.003	0.003	0.003	0.003	0.003
-0.683	0.317	0.009	0.009	0.005	0.010	0.010	1.096	2.096	0.003	0.003	0.003	0.003	0.004
-0.648	0.352	0.010	0.011	0.008	0.009	0.012	1.132	2.132	0.003	0.003	0.003	0.003	0.003
-0.612	0.388	0.010	0.012	0.007	0.010	0.012	1.167	2.167	0.003	0.003	0.002	0.003	0.003
-0.577	0.423	0.012	0.012	0.011	0.013	0.013	1.203	2.203	0.002	0.002	0.002	0.002	0.003
-0.541	0.459	0.013	0.014	0.010	0.014	0.016	1.238	2.238	0.002	0.002	0.002	0.002	0.002
-0.505	0.495	0.015	0.017	0.011	0.014	0.016	1.274	2.274	0.002	0.002	0.002	0.002	0.003
-0.470	0.530	0.016	0.017	0.012	0.017	0.019	1.309	2.309	0.002	0.003	0.001	0.002	0.003
-0.434	0.566	0.017	0.019	0.012	0.017	0.017	1.345	2.345	0.002	0.002	0.001	0.002	0.002
-0.399	0.601	0.019	0.022	0.015	0.020	0.018	1.381	2.381	0.002	0.002	0.002	0.001	0.002
-0.363	0.637	0.020	0.023	0.017	0.020	0.017	1.416	2.416	0.002	0.002	0.002	0.001	0.003
-0.327	0.673	0.022	0.024	0.019	0.022	0.022	1.452	2.452	0.002	0.002	0.001	0.002	0.002
-0.292	0.708	0.024	0.026	0.022	0.023	0.023	1.487	2.487	0.001	0.001	0.001	0.001	0.001
-0.256	0.744	0.025	0.024	0.024	0.025	0.025	1.523	2.523	0.001	0.002	0.001	0.002	0.001
-0.221	0.779	0.026	0.029	0.025	0.026	0.025	1.559	2.559	0.001	0.001	0.001	0.001	0.002
-0.185	0.815	0.030	0.030	0.029	0.029	0.030	1.594	2.594	0.001	0.001	0.001	0.002	0.002
-0.150	0.850	0.030	0.032	0.029	0.029	0.031	1.630	2.630	0.001	0.001	0.001	0.002	0.002
-0.114	0.886	0.031	0.031	0.032	0.031	0.028	1.665	2.665	0.001	0.001	0.001	0.001	0.001
-0.078	0.922	0.032	0.031	0.035	0.032	0.028	1.701	2.701	0.001	0.001	0.001	0.001	0.001
-0.043	0.957	0.036	0.037	0.039	0.034	0.030	1.737	2.737	0.001	0.001	0.001	0.001	0.001
-0.007	0.993	0.035	0.037	0.037	0.034	0.031	1.772	2.772	0.001	0.001	0.001	0.001	0.001
0.028	1.028	0.035	0.034	0.042	0.033	0.032	1.808	2.808	0.001	0.001	0.001	0.001	0.001
0.064	1.064	0.036	0.035	0.042	0.036	0.031	1.843	2.843	0.001	0.001	0.001	0.001	0.001
0.100	1.100	0.035	0.033	0.040	0.035	0.035	1.879	2.879	0.001	0.001	0.001	0.001	0.001
0.135	1.135	0.034	0.031	0.038	0.035	0.032	1.914	2.914	0.001	0.001	0.001	0.000	0.001
0.171	1.171	0.031	0.030	0.038	0.031	0.026	1.950	2.950	0.001	0.001	0.000	0.001	0.001
0.206	1.206	0.030	0.028	0.038	0.028	0.029	1.986	2.986	0.001	0.001	0.000	0.001	0.001
0.242	1.242	0.027	0.026	0.033	0.027	0.024	2.021	3.021	0.001	0.001	0.000	0.001	0.000
0.277	1.277	0.026	0.024	0.030	0.025	0.026	2.057	3.057	0.000	0.000	0.000	0.000	0.001
0.313	1.313	0.023	0.021	0.028	0.024	0.020	2.092	3.092	0.001	0.000	0.000	0.001	0.001
0.349	1.349	0.021	0.019	0.025	0.021	0.020	2.128	3.128	0.001	0.001	0.000	0.001	0.001
0.384	1.384	0.020	0.017	0.023	0.021	0.019	2.164	3.164	0.001	0.001	0.001	0.001	0.000
0.420	1.420	0.018	0.017	0.023	0.018	0.015	2.199	3.199	0.000	0.000	0.000	0.000	0.001
0.455	1.455	0.016	0.015	0.019	0.015	0.015	2.235	3.235	0.001	0.001	0.000	0.001	0.001
0.491	1.491	0.013	0.013	0.014	0.013	0.014	2.270	3.270	0.000	0.000	0.000	0.001	0.001
0.527	1.527	0.013	0.011	0.014	0.012	0.013	2.306	3.306	0.001	0.000	0.000	0.001	0.001
0.562	1.562	0.011	0.012	0.011	0.011	0.010	2.341	3.341	0.000	0.000	0.000	0.001	0.000
0.598	1.598	0.010	0.011	0.009	0.011	0.009	2.377	3.377	0.001	0.000	0.000	0.001	0.001
0.633	1.633	0.009	0.009	0.009	0.009	0.009	2.413	3.413	0.000	0.000	0.000	0.000	0.001
0.669	1.669	0.008	0.008	0.009	0.008	0.008	2.448	3.448	0.000	0.000	0.000	0.000	0.000
0.705	1.705	0.008	0.008	0.007	0.008	0.009	2.484	3.484	0.000	0.000	0.000	0.000	0.001
0.740	1.740	0.007	0.008	0.006	0.006	0.008	2.519	3.519	0.000	0.000	0.000	0.000	0.000
0.776	1.776	0.006	0.006	0.006	0.006	0.006	2.555	3.555	0.000	0.000	0.000	0.000	0.001
0.811	1.811	0.006	0.005	0.006	0.006	0.006	2.591	3.591	0.000	0.000	0.000	0.001	0.000
0.847	1.847	0.005	0.004	0.005	0.006	0.007	2.626	3.626	0.000	0.000	0.000	0.001	0.000
0.882	1.882	0.005	0.005	0.004	0.005	0.005	2.662	3.662	0.000	0.000	0.000	0.000	0.000
0.918	1.918	0.004	0.004	0.002	0.004	0.005	2.680	3.680	0.010	0.010	0.007	0.012	0.014

Table E2. Three-Year Return Distribution

Three-year return distribution													
		All	Zero	Old	Mid	New			All	Zero	Old	Mid	New
R	W	P	P	P	P	P	R	W	P	P	P	P	P
-0.871	0.129	0.033	0.035	0.020	0.034	0.044	2.636	3.636	0.002	0.002	0.002	0.003	0.002
-0.835	0.165	0.021	0.023	0.016	0.019	0.028	2.707	3.707	0.002	0.002	0.002	0.002	0.003
-0.765	0.235	0.023	0.023	0.016	0.025	0.028	2.778	3.778	0.002	0.002	0.002	0.002	0.002
-0.694	0.306	0.026	0.026	0.019	0.029	0.029	2.849	3.849	0.002	0.002	0.001	0.002	0.002
-0.623	0.377	0.027	0.028	0.019	0.029	0.032	2.919	3.919	0.002	0.002	0.001	0.002	0.002
-0.552	0.448	0.029	0.030	0.023	0.031	0.032	2.990	3.990	0.001	0.002	0.001	0.001	0.002
-0.481	0.519	0.030	0.031	0.026	0.031	0.034	3.061	4.061	0.002	0.002	0.001	0.002	0.002
-0.410	0.590	0.032	0.032	0.029	0.032	0.036	3.132	4.132	0.001	0.001	0.001	0.001	0.001
-0.339	0.661	0.034	0.035	0.030	0.035	0.035	3.203	4.203	0.001	0.002	0.001	0.002	0.002
-0.269	0.731	0.035	0.034	0.034	0.034	0.038	3.274	4.274	0.001	0.001	0.001	0.002	0.001
-0.198	0.802	0.037	0.038	0.037	0.037	0.036	3.344	4.344	0.001	0.001	0.001	0.001	0.001
-0.127	0.873	0.038	0.038	0.040	0.037	0.035	3.415	4.415	0.001	0.001	0.001	0.001	0.001
-0.056	0.944	0.036	0.037	0.038	0.035	0.032	3.486	4.486	0.001	0.001	0.001	0.001	0.001
0.015	1.015	0.038	0.038	0.039	0.041	0.035	3.557	4.557	0.001	0.001	0.001	0.001	0.001
0.086	1.086	0.037	0.037	0.043	0.036	0.031	3.628	4.628	0.001	0.002	0.001	0.001	0.001
0.156	1.156	0.037	0.038	0.042	0.036	0.032	3.699	4.699	0.001	0.001	0.001	0.001	0.001
0.227	1.227	0.037	0.035	0.045	0.035	0.032	3.770	4.770	0.001	0.001	0.001	0.001	0.001
0.298	1.298	0.034	0.035	0.037	0.031	0.028	3.840	4.840	0.001	0.001	0.001	0.001	0.001
0.369	1.369	0.033	0.033	0.040	0.031	0.029	3.911	4.911	0.001	0.001	0.001	0.000	0.001
0.440	1.440	0.032	0.031	0.039	0.031	0.025	3.982	4.982	0.001	0.001	0.000	0.001	0.001
0.511	1.511	0.030	0.027	0.039	0.029	0.028	4.053	5.053	0.001	0.001	0.000	0.001	0.001
0.582	1.582	0.026	0.026	0.030	0.025	0.025	4.124	5.124	0.001	0.001	0.000	0.001	0.001
0.652	1.652	0.025	0.023	0.031	0.024	0.022	4.195	5.195	0.001	0.001	0.000	0.000	0.001
0.723	1.723	0.022	0.022	0.026	0.023	0.016	4.265	5.265	0.001	0.001	0.000	0.001	0.001
0.794	1.794	0.021	0.018	0.026	0.020	0.019	4.336	5.336	0.001	0.000	0.000	0.001	0.001
0.865	1.865	0.019	0.018	0.022	0.018	0.018	4.407	5.407	0.001	0.001	0.000	0.000	0.001
0.936	1.936	0.016	0.016	0.019	0.016	0.014	4.478	5.478	0.000	0.001	0.000	0.001	0.000
1.007	2.007	0.014	0.016	0.014	0.013	0.011	4.549	5.549	0.001	0.001	0.000	0.001	0.000
1.077	2.077	0.014	0.012	0.015	0.016	0.013	4.620	5.620	0.000	0.000	0.000	0.001	0.001
1.148	2.148	0.012	0.012	0.013	0.011	0.012	4.691	5.691	0.000	0.001	0.000	0.000	0.000
1.219	2.219	0.011	0.011	0.013	0.009	0.011	4.761	5.761	0.000	0.001	0.000	0.000	0.000
1.290	2.290	0.010	0.009	0.012	0.010	0.012	4.832	5.832	0.000	0.000	0.000	0.000	0.000
1.361	2.361	0.009	0.008	0.010	0.009	0.008	4.903	5.903	0.000	0.000	0.000	0.000	0.000
1.432	2.432	0.008	0.007	0.007	0.009	0.008	4.974	5.974	0.000	0.001	0.000	0.000	0.001
1.502	2.502	0.008	0.008	0.007	0.009	0.008	5.045	6.045	0.000	0.001	0.000	0.000	0.001
1.573	2.573	0.006	0.006	0.007	0.006	0.007	5.116	6.116	0.000	0.000	0.000	0.001	0.001
1.644	2.644	0.007	0.007	0.007	0.005	0.008	5.186	6.186	0.000	0.000	0.000	0.000	0.001
1.715	2.715	0.006	0.007	0.004	0.004	0.007	5.257	6.257	0.000	0.001	0.000	0.000	0.001
1.786	2.786	0.005	0.005	0.006	0.005	0.005	5.328	6.328	0.000	0.000	0.000	0.000	0.000
1.857	2.857	0.005	0.005	0.005	0.005	0.007	5.399	6.399	0.000	0.000	0.000	0.000	0.000
1.928	2.928	0.005	0.004	0.004	0.005	0.004	5.470	6.470	0.000	0.000	0.000	0.001	0.000
1.998	2.998	0.004	0.004	0.004	0.004	0.005	5.541	6.541	0.000	0.000	0.000	0.000	0.001
2.069	3.069	0.004	0.004	0.005	0.004	0.004	5.612	6.612	0.000	0.001	0.000	0.000	0.000
2.140	3.140	0.003	0.004	0.002	0.004	0.004	5.682	6.682	0.000	0.001	0.000	0.000	0.000
2.211	3.211	0.003	0.003	0.003	0.002	0.003	5.753	6.753	0.000	0.000	0.000	0.000	0.001
2.282	3.282	0.003	0.003	0.002	0.003	0.002	5.824	6.824	0.000	0.000	0.000	0.000	0.000
2.353	3.353	0.003	0.003	0.002	0.003	0.004	5.895	6.895	0.000	0.001	0.000	0.000	0.001
2.423	3.423	0.002	0.002	0.002	0.003	0.003	5.966	6.966	0.000	0.000	0.000	0.001	0.000
2.494	3.494	0.002	0.002	0.002	0.003	0.003	6.037	7.037	0.000	0.000	0.000	0.000	0.000
2.565	3.565	0.002	0.002	0.001	0.002	0.003	6.072	7.072	0.010	0.009	0.006	0.013	0.014

3 The Stock Market Valuation of Corporate Social Responsibility

Modern-day corporations have increasingly embraced corporate social responsibility (CSR). According to a 2019 survey of 350 business leaders conducted by Deloitte Global and Forbes Insights, 93 percent of executives view their companies as stewards of society.[17] Additionally, the scope of CSR has expanded beyond simply donating money to charitable organizations to include addressing a variety of societal challenges, such as environmental, educational, and diversity issues, among others.

As Environmental, Social, and Governance (ESG) criteria become increasingly influential, there is a growing debate regarding the effect of CSR on firm value. Critics argue that CSR can be a distraction from business objectives and erodes shareholder value, consistent with the classical view of Friedman (1970). On the other hand, proponents consider CSR to be an essential part of corporate strategy for long-term success, as suggested by Freeman (2010). However, the debate is often complicated due to the limited data on how firms incorporate CSR principles into their practices. For instance, ESG ratings or scores, a primary source for CSR research, lack details on corporate actions directly advancing CSR objectives, thereby clouding the ability to independently assess their impact. Consequently, both supporters and detractors of CSR initiatives frequently rely more on heuristics and biases than on empirical evidence for their conclusions.

To address this issue, I compile a dataset from corporate press releases that detail companies' CSR activities. CSR activities or initiatives are projects, programs, or events conducted by companies to address social and environmental issues beyond their legal obligations. For example, in 2018, Amazon invested in the Closed Loop Fund to improve recycling infrastructure in the US. Their objective was to divert 1 million tons of recyclable material from landfills and cut CO2 emissions by 2 million metric tons by 2028. Unlike ESG scores or generic ESG news, this dataset more accurately reflects the deliberate CSR activities of firms, providing a place for a clearer understanding of how these efforts are perceived and valued in the market.

This paper studies how CSR initiatives influence firm value, as reflected in market reactions to them. Like other corporate projects, the value of a CSR activity is assessed through its incremental cash flows. These cash flows are determined by factoring in opportunity costs such as forgone investments (Bénabou and Tirole (2010))

[17]"The Rise of the Socially Responsible Business" 2019. Deloitte Touche Tohmatsu Limited. https://deloitte.wsj.com/articles/the-rise-of-the-socially-responsible-business-01548381736

and the benefits of positive externalities, including potentially lower borrowing costs (Goss and Roberts (2011), Höck et al. (2020), Flammer (2021)), reduced required return on equity (Hong and Kacperczyk (2009), El Ghoul et al. (2011), Dhaliwal et al. (2011), Flammer (2013), Flammer (2015), Riedl and Smeets (2017), Ardia et al. (2020), Pástor, Stambaugh, and Taylor (2021)), increased consumer willingness to pay premiums (Ha-Brookshire and Norum (2011), Anselmsson, Vestman Bondesson, and Johansson (2014)), favorable media coverage (Dhaliwal et al. (2011), Cahan et al. (2015)) and decreased litigation risks (Badawi and Partnoy (2022)). Given that CSR activities themselves are not a significant source of direct profits, comprehending their value necessitates an examination of CSR's influence on the indirect elements of incremental cash flows. Specifically, for CSR to be valuable, it must generate positive externalities on current operations, potentially exceeding opportunity costs and negative externalities on operations. Therefore, this study focuses on a factor that could influence the positive externalities of CSR initiatives on ongoing operations: the public's demand for CSR.

Specifically, I suggest that CSR initiatives can increase firm value when they are executed with an understanding of the public's needs. If the public perceives a pressing need for corporate intervention in critical issues, a wide range of people might engage in activities that ultimately provide incentives to firms that address these issues. For instance, in line with the literature, both investors and consumers, who constitute part of the public, may be willing to pay a premium to support companies that address these issues. Conversely, interest groups or regulatory bodies may penalize companies that overlook environmental or societal issues, potentially leading to litigation. Without such incentives and penalties, companies may remain passive, making the cost of addressing social issues prohibitively high for individuals.

However, the public's demand for CSR is context-dependent and varies in intensity. In this regard, I investigate two crucial factors that capture the variation in public needs for CSR. The first factor is the level of public concern about a social issue, with more critical issues garnering greater support and less critical ones receiving less. Secondly, the comparative advantage companies hold in tackling an issue plays a role. There are issues that companies may have a distinct edge over individuals or even governments. In such cases, CSR initiatives that surpass what the average individual can achieve might be more highly valued. Initiatives that are easily replicable by individuals do not yield extra benefits for them (Bénabou and Tirole (2010)).

I begin the paper by introducing the primary data for this study, a set of press releases on CSR activities that firms voluntarily release. I select the largest 1,000 companies each year and track their press releases from 2006 to 2020 from corporate websites and Factiva.[18] This process results in a sample free of survivorship bias. Then, I develop a multi-label classification deep learning model to identify CSR news releases. The model achieves 97% recall on the test set. After classifying press releases with the model, I further eliminate CSR news releases concerning ESG index membership, ESG ratings, ESG awards, CEOs' statements, and other content not directly related to company actions. Finally, I manually label CSR news releases with the social issues that a CSR activity addresses and the methods of implementation, such as projects, operational changes, and donations, as described in the news. The resulting dataset contains 23,698 news releases on CSR initiatives, with 54% of them addressing environmental, inclusion, poverty, and education issues in the U.S.

With the news releases on CSR activities, I analyze market reactions to them. I calculate the four-factor adjusted cumulative abnormal returns based on Carhart (1997), over a three-day window that includes the day before, the day of, and the day after a CSR activity news release. A change in stock price within this narrow timeframe reflects the value of the CSR activity. If the action is seen as value-enhancing, the stock price is expected to increase. The collection of abnormal returns surrounding those news releases provides a valuable space for identifying factors that systematically affect the valuation of CSR initiatives.

On average, without any risk adjustment, CSR activities generate positive three-day cumulative returns. However, when adjusted for risk, the average returns level out to zero. Both raw returns and risk-adjusted returns exhibit considerable variation, with a standard deviation of about 0.03. Moreover, there is a symmetry of market reactions around the median, prompting an investigation into the sources of this variation.

To examine the relationship between the level of public concern about an issue and related corporate actions, I use the frequency of discussions about an issue in newspapers as a proxy. This method aligns with previous studies that utilize news articles to gauge the public's level of concern regarding a particular social matter.[19]

[18] For companies with updated websites that often do not archive press releases from before certain years, I turn to Factiva to gather these documents. Additionally, by 2021 - the year in which the news releases were compiled - some firms had either ceased operations or were acquired by other companies. For these entities as well, I depend on Factiva to retrieve their news releases as their websites are not available. When collecting press releases from Factiva, I ensure that the company's name is included in the source section.

[19] Engle et al. (2020) build an index that captures the attention to climate change in the Wall Street Journal. Ardia et al. (2020) use data from various news outlets to capture concerns over climate change.

I curate articles from the opinion, letter, interview, comment, column, and editorial sections of the Wall Street Journal, New York Times, and USA Today.

I specifically focus on the level of public concern regarding the environment, inclusion, poverty, and education in the U.S., as these issues represent the majority of social issues addressed by CSR activities in the sample. For each of the four issues, I handpick representative articles. Then, using a Natural Language Processing (NLP) algorithm that is based on document similarity, I identify articles deemed to address a particular social problem. Next, I obtain monthly counts of articles for each issue, sorted by news outlet. These counts are then normalized within each outlet and averaged across the three outlets on a monthly basis. After performing a monotone transformation, I have a monthly index that represents public concern for each of the four social issues.

I regress market reactions on the degree of public concern, averaged over the preceding three or six months, relating to the specific social issue targeted by a CSR activity. For instance, if a company launches a state-wide coding program for public high schools, the related CSR news is assigned the public concern about education in the U.S. over the past three or six months. Incorporating issue-fixed effects to focus on the temporal variation in the level of public concern, I find it positively influences the market's reaction to that program. A one standard deviation increase in public concern gauged over the prior three months, leads to a market reaction that is 0.061 percentage points higher. When using the metric based on the past six months, this market reaction rises by 0.058 percentage points. The findings support the hypothesis that CSR can positively affect firm value when tackling social issues that weigh heavily on the public, given its potential to garner extensive backing from stakeholders who can influence the firm.

Next, I explore whether investors assign greater value to CSR initiatives when they address the issue more effectively than individuals can. Building on the concept of delegated philanthropy described by Bénabou and Tirole (2010), I posit that while individuals can support various philanthropic causes through charitable organizations, there are instances where direct corporate action provides more effective solutions.

I categorize environmental and inclusion challenges as areas where corporate responses, which go beyond mere monetary contributions, can be more effective than individual efforts. It is because companies are intrinsically linked to these issues. Environmental impacts frequently stem directly from their operations, and they wield significant influence over workforce policies and community engagement through products and services. Those corporations, with their substantial and relevant resources, are ideally positioned to address them effectively.

To distinguish CSR activities that individuals might also be able to undertake, I classify CSR activities into two categories: corporate giving and other forms of actions, which include operational changes, shifts in firm policies, regional or nationwide projects, etc. I then examine how the method of addressing social issues interacts with the type of social issues to generate different market reactions. I find when the social issue being addressed is related to environmental or inclusion concerns, CSR initiatives other than corporate giving result in a 0.3 percentage point higher market reaction on average. The finding aligns with the notion that companies that effectively address social issues can garner support from the public, potentially leading to a positive impact on firm value.

I further explore whether the two factors continue to influence market reactions, even when considering firm characteristics known to affect CSR. I first examine the profitability. If the value of CSR activities is merely determined by the opportunity costs associated with them, then including profitability metrics could eclipse previous findings. This is because more profitable firms face lower opportunity costs due to financial slack. I evaluate profitability by the past two-year return on assets (ROAs) and recent earnings surprises. An increase of one standard deviation in the ROAs over the past two years corresponds to a 0.14 percentage point increase in market reactions. Similarly, a standard deviation increase in recent EPS surprises results in a 0.15 percentage point rise in CSR news release returns, highlighting investors' awareness of the recent financial performance of a firm and CSR outlays. However, even accounting for these variables, the effect of public concern and the CSR approach persists.

Secondly, I investigate the effect of the governance of a firm. CSR actions could be influenced by agency problems. Research by Cheng, Hong, and Shue (2013) and Masulis and Reza (2014) indicates that corporate donations may lead to a misuse of resources and a reduction in firm value when they align with the self-interests of the CEO. Similarly, Di Giuli and Kostovetsky (2014) find that high ESG ratings sometimes mirror the political biases of company executives and can predict future stock return declines and lower ROAs. However, others show that governance can shape the adoption of CSR policies and positively relate to the valuation of CSR initiatives. For instance, Ferrell, Liang, and Renneboog (2016) show that well-governed firms, with minimal agency issues, engage more in CSR, and CSR can reduce managerial entrenchment costs.

If CSR is mainly associated with corporate governance, adding related metrics might mask the main findings. I include board structure variables and find firms with a CEO serving the role of chairman see a 0.3 percentage point decrease in market response to CSR programs. This is consistent with the prior studies indicating that CSR influenced by CEO preferences may carry higher agency costs. However, even after accounting for board characteristics, the effects of public concern and the methods of CSR implementations remain statistically significant.

With insights into how public concern and the method of CSR implementations influence market reactions, I examine whether companies integrate these elements into their CSR strategies. From monthly panel regressions, I find that public concern over the past one to twelve months pushes firms toward CSR actions on related issues. For example, when it comes to environmental issues, a one standard deviation increase in one-year public concern boosts the likelihood of a firm launching environment-CSR initiatives by 0.188 percentage points, which translates into a 15.67% increased probability from the unconditional likelihood.

Furthermore, companies choose methods to address issues strategically. In regression analyses with a sample of CSR activities, as public concern intensifies regarding environmental and diversity matters, companies opt for approaches that go beyond just corporate donations. For poverty and education concerns they are more likely to choose corporate giving.

This study is the first systematic investigation into corporate activities pursued explicitly under the banner of CSR causes. Through an exhaustive investigation of CSR activities across 1430 large public firms from 2006 to 2020, I show not all CSR initiatives enhance value. Initiatives that increase value are those that align with public discourse on environmental and social issues and exhibit effectiveness in addressing these concerns.

The rise in CSR activities among firms aligns with the growing perspective that CSR has evolved into a crucial business practice. In 2020, over half of the firms in my sample reported involvement in more than one CSR initiative, a stark increase from the less than 20% recorded in 2006. This is a paradigm shift from the conventional views posited by Friedman (1970).

This shift can be attributed to various catalysts, including changes in investor preferences (Hartzmark and Sussman (2019)), the primary types of assets firms leverage (Edmans (2011)), and the evolving definitions of corporate responsibility (Hart and Zingales (2017)), among others. I introduce a perspective that, in light of the complexity of social and environmental challenges and the unique positions of firms, there are times when a good portion of the population views corporate participation as vital for tackling these issues. As a result, those in need of corporate actions mete out both incentives and penalties to induce profit-maximizing firms to play an active role in managing these challenges. Hence, strategically timed CSR initiatives and well-chosen methods can provide an opportunity for a company to enhance its value.

The paper sheds light on the mixed findings in the literature regarding CSR's impact on firm value and offers a discussion on it. Margolis, Elfenbein, and Walsh (2009) conduct a meta-analysis with more than 250 CSR papers and find there is little evidence supporting a statistically significant relation between CSR and firm performance. Also, research into market responses to broad positive ESG news, which covers news beyond what companies explicitly report to undertake, yields mixed findings. According to Krüger (2015), positive generic ESG news leads to negative market reactions. Capelle-Blancard and Petit (2017) suggests that companies gain nothing from ordinary positive ESG news. On the other hand, positive general ESG news receives positive reactions according to Dimson, Karakaş, and Li (2015) and Serafeim and Yoon (2021).

By focusing on a collection of press releases detailing CSR activities, I find that the market reactions to these activities are, on average, neutral. I discuss four reasons for this observation, drawing on findings presented in the paper as well as those detailed in the Online Appendix. Firstly, virtue-signaling CSR activities may fail to enhance firm value, evidenced by the fact that corporate donations elicit significantly negative or lukewarm market reactions, depending on the specific issues addressed. Secondly, while profitability is a key factor influencing market response, there is a lack of evidence it affects a company's engagement in CSR initiatives. This indicates that firms might be overlooking the significance of profitability in the formulation of CSR goals. Thirdly, the CEO's power within the board - another negative predictor of the market value of CSR - relates positively to philanthropic CSR engagement. This suggests that CSR activities might reflect other decisions by CEOs that stray from the pursuit of maximizing firm value. Finally, there could be the presence of peer effects in the adoption of CSR activities, leading companies to engage in CSR initiatives without a carefully considered strategy for enhancing firm value.

The remainder of the paper is structured as follows: The first section presents the data. Section 2 examines the two factors linked to public demand for CSR that influence market reactions to CSR activities. Section 3 explores how companies' CSR decision-making aligns with the factors explored in Section 2. Section 4 discusses the observation of average zero returns surrounding CSR activity news release days. The paper concludes with Section 5.

3.1 Data

This section explains my primary data, corporate press releases, and how I collect and classify them. I first gather all corporate press releases whose sources are corporations themselves from corporate websites and Factiva. Since the coverage of press releases of Faciva increased dramatically in 2006, I restrict my sample to a period between 2006 and 2020. I select U.S. firms whose end-of-fiscal-year market capitalization ranks within the top 1,000 among all U.S. firms in a given year from 2006 to 2020 and exclude financial and utility firms. I found 1,502 such firms. Among them, I investigated 1,430 firms and collected press releases for these firms over the entire sample period. This process resulted in a survivorship bias-free sample. More specific details of the sample selection and descriptive statistics are explained in Appendix A.

After collecting news releases, I classify them into a predetermined set of categories. Specifically, a news release is classified into one or more than one of the categories: (1) earnings/performance related news, (2) M&A related news, (3) CSR news, (4) news on executives, (5) news on financing activities such as equity issuance, debt issuance, or retirement of existing debt, (6) news on directors, (7) stock repurchase news, (8) dividend news, and (9) other business-related news.

To classify the collected press releases, I create a deep learning model to facilitate the classification. I use transfer learning in natural language processing (NLP). Natural language processing is a tool to transform text into quantifiable numbers based on features of the text and transfer learning in NLP leverages prior knowledge from prior work. Specifically, I use a Bidirectional Encoder Representations from Transformers (BERT) (2018) model introduced by Devlin et al. (2019) to translate an article into a vector (an embedding) of 768 dimensions. BERT is a machine learning technique trained on a vast corpus extracted from the BooksCorpus with 800M words and English Wikipedia with 2,500M words and is widely used as it is known to generate accurate word representations. In this setting, the transfer learning using a pre-trained BERT model helps me exploit knowledge the pre-trained model gained from the vast corpus texts.

After training a multi-label classification model with transfer learning, I input all press releases into the model to get predictions. I inspect the press releases classified as CSR news by the model and eliminate any news releases that are misclassified. I further eliminate CSR news releases that do not report actual CSR activities. For example, I remove news releases on the membership of an ESG index, any ESG-related awards, CEOs' statements, and other information that do not contain actions taken by the company. The complete steps to train the model and the model performance on the test set are provided in Appendix B. If the same CSR activity is reported multiple times by different news outlets, I keep the earliest news release.

Table 1 presents the total number of firms and the total number of CSR activities reported annually. The number of active firms varies each year, introducing a fluctuation in the count, despite the analysis encompassing 1,430 firms from 2006 to 2020. There is a clear upward trend in the number of reported activities over time, starting from 689 in 2006 and rising to 2,101 by 2020. In terms of proportions, less than 20% of firms engaged in CSR in 2006, while more than 50% of firms participated in some form of CSR in 2020. Shifting the focus to the descriptive statistics on annual CSR activities per firm for those that reported at least one CSR activity, the average number of activities typically hovers around four. The median remains constant at two. Excluding maximum values, the distribution of activities per firm is quite narrow, generally varying from one to four, based on the 25-75 percentiles. Walmart is the company that frequently reports the maximum number of activities in a year in the sample, actively engaging in various types of CSR activities.

Table 2 illustrates that numerous companies implement CSR in a variety of forms. Poverty emerges as the primary concern addressed by firms, representing 17.1% of all CSR activities in the sample. This is followed by educational initiatives, accounting for 14.8%. Environmental CSR efforts make up 13.6% of the total CSR activities reported, with initiatives related to inclusion and diversity comprising 9.38%.

The better part of the paper investigates market reactions to corporate news. Market reactions are defined as daily cumulative abnormal returns. I use Carhart (1997) four-factor model to compute risk-adjusted daily returns. Daily returns are obtained from the Center for Research in Security Prices (CRSP) daily return file. Daily four factors are obtained from Ken French's website. I estimate alpha and betas using estimation periods of 365 days ending 50 days before the event date while eliminating estimates that are computed with less than 200 daily returns. Specifically, these parameters are estimated by regressing excess daily returns on daily factors. I compute cumulative abnormal returns around an event day with a window of 3 days as the summation of daily risk-adjusted returns on the event day, the day before the event day, and the day after the event day.

Table 2 provides summary statistics of market reactions to CSR activities. When measured by raw returns, without any risk adjustment, CSR activities seem to generate, on average, positive returns. However, risk-adjusted returns average out to zero. This neutral market reaction to CSR activities will be further addressed in the discussion section of the paper. Examining other summary statistics, the market reactions tend to be symmetric around the median, indicating no evident skewness. Additionally, the reactions display significant variation, with a standard deviation of approximately 0.03—this is considerable, given that the returns are measured over a three-day period. The fact that the 75th percentile is around 0.011 suggests that there are instances where CSR activities increase firm value by 1.1%, while the 25th percentile being -0.011 indicates that the likelihood of CSR activities decreasing firm value by 1.1% is nearly the same. These observations prompt an investigation into which factors systematically influence the value of CSR activities either positively or negatively.

I gather accounting variables from Compustat's North America Fundamentals Annual database and EPS variables from the I/B/E/S database. Board characteristics are sourced from BoardEx.

3.1.1 Measures of Public Concern

To measure public concern over a particular social issue, I collect columns, editorials, letters, opinions, and interviews published from 1996 to 2020 in the New York Times, Wall Street Journal, and USA Today.[20] According to communications studies, journalism and viewers' demand for news influence each other. Journalism affects viewers by highlighting particular events, choosing narratives, and the frequency of publications of news. On the other hand, readers' demand for specific information or their opinions also affects journalism. A number of studies show that news content is determined by who is interested in it and its value to advertisers (Hamilton (2003)). Regardless of which mechanism dominates the other, what is clear is that the number of articles related to a certain issue and readers' concern over the issue are likely to be highly correlated.[21]

Having more than 500,000 articles collected from the news outlets, my goal is to identify articles that cover social issues and count them to construct public concern measures. In particular, I focus on four social issues: environmental issues, poverty issues, education, and diversity-related issues in the U.S. since the news releases reporting CSR activities addressing these social issues make up more than 54% of all news releases on CSR activities in the sample.

[20] I collected columns, editorial, letters, opinions, comments, and interviews rather than entire news sections because citizens' opinions are likely to be reflected in these six sections.

[21] There are several papers that measure investors' concern or attention from news media. Engle et al. (2020) build an index that captures the attention to climate change in the Wall Street Journal. Ardia et al. (2020) use news data to capture concerns over climate change, and Pástor, Stambaugh, and Taylor (2021) use the same measure.

To make the search process tractable, I rely on a natural language processing (NLP) method. I first select more than forty representative articles that discuss each social issue. Next, I use a Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. (2019)) model to translate collected articles into a vector of 768 dimensions. In particular, I use Sentence-BERT since it produces better embeddings suitable for computing similarity measures (Reimers and Gurevych (2019)). More specifically, I feed the title, first three sentences, and last three sentences of an article into the NLP model to get embeddings for each sentence and average them to produce the embedding for the article.

To determine whether an article covers a specific social issue, I compute the cosine similarity between the embedding of the article and the embedding of each of the forty representative articles of the social issue. If the cosine similarity between the article and any one of those forty representative articles is greater than 0.93, I classify the article as concerned with the social issue.

I count the number of articles that are predicted to be related to a specific social issue by month while tallying separately for each news outlet to account for heterogeneity across the outlets. Following Baker, Bloom, and Davis (2016) and Ardia et al. (2020), I divide the number by the standard deviation of it in each source before aggregation. Then, I average the numbers across news outlets each month. The process results in monthly time series observations for each social issue ($\{x_{t,p}\}$, $t \in \{Jan2006, Feb2006, \ldots, Dec2020\}$, $p \in \{Environment, Inclusion, Education, Poverty\}$). I compute the moving average of the monthly observations over a horizon of three months, six months, or twelve months in the past ($Concern_{t-k,t,p} = \frac{1}{k}\sum_{t-k+1}^{t} x_{\tau,p}$, $k \in \{3, 6, 12\}$).

In some model specifications, I employ a monotone-transformed metric of public concern to enhance the explanatory power of the variable. This methodology is permissible given that there are no parametric restrictions on the measurement of public concern. Following Ardia et al. (2020), I apply an increasing concave function to normalize the variable. I take the log of the moving averages (plus one), $log(1 + \frac{1}{k}\sum_{t-k+1}^{t} x_{\tau,p})$. Later, the values of this variable are assigned to CSR activities based on the issues they address and the dates on which these activities are reported.

Table 3 presents the summary statistics for the measures of public concern, $Concern_{t-k,t,p}$. These measures display significant temporal variations. Although public concern for environmental issues, on average, exceeds that for other social issues, there is significant variation within the public concern related to each specific issue.

Figure 1 presents the public concern indices and the number of CSR activities from 2006 to 2020. The public concern index for environmental issues shows a discernible upward trend in recent years, indicating a growing public awareness and concern for environmental issues. The public concern for diversity appears relatively stable over the sample period, with occasional spikes. However, there is a sharp increase towards the end, which aligns with the period following the George Floyd protests that ignited the Black Lives Matter movement. As for public concern for education and poverty, despite visible fluctuations, there does not appear to be a clear long-term increasing or decreasing trend.

Examining the number of CSR activities, I find there is a rising trend in environmental CSR activities over time. They start modestly but have gained momentum in recent years. The spikes observed every April can be attributed to companies' tendency to report their environmental initiatives on Earth Day. Initially, diversity-related CSR initiatives are relatively low compared to other issues. However, a significant increase is evident in later years, potentially correlating with the heightened societal focus on diversity and inclusion, particularly in the context of the Black Lives Matter movement. Education-focused CSR actions characterize the early days of CSR activities and maintain a stable presence over time. However, their prevalence has decreased in more recent years. Similarly, CSR efforts addressing poverty are consistent throughout the years, with a more pronounced activity towards the end of the period. This could potentially be a response to the economic impacts of the Covid-19 crisis.

3.2 Market Reactions to CSR Activities

In this section, I investigate whether factors associated with public demand for CSR influence the market valuation of CSR initiatives. Firstly, I consider the degree of public concern tied to a particular social issue that a company addresses. Secondly, I delve into the specific methods employed by firms in addressing these social issues.

3.2.1 Public Concern and Market Reaction to CSR Activities

In this section, I examine the effect of the level of public concern on the market reaction to CSR initiatives. The hypothesis posits that companies receive a more favorable market response when they tackle social issues that are of paramount public concern, as such actions are likely to garner widespread support for the firm from the general public, which could positively affect the firm.

The sample contains news releases on CSR activities that can be linked to one of the four social issues. Contingent on a specific social issue and a month in which a CSR activity is reported, I generate a CSR activity-specific measure of public concern, denoted as $CSR\_Concern(M)$, where M represents the period in months over which the monthly public concern index is averaged.

To illustrate, consider a press release announcing a firm's commitment to reduce greenhouse gas emissions in month t. The corresponding $CSR\_Concern(3)$ for this announcement is calculated as $log(1 + \frac{1}{3}\sum_{\tau=t-2}^{t} x_{\tau,Environment})$. Likewise, $CSR\_Concern(6)$ for the press release is $log(1 + \frac{1}{6}\sum_{\tau=t-5}^{t} x_{\tau,Environment})$. Similarly, for a press release reporting a firm's initiative to host a nationwide scientific competition for K-12 students published in month t, the $CSR\_Concern(3)$ is given by $log(1 + \frac{1}{3}\sum_{\tau=t-2}^{t} x_{\tau,Education})$ and $CSR\_Concern(6)$ for the press release is $log(1 + \frac{1}{6}\sum_{\tau=t-5}^{t} x_{\tau,Education})$.

With the CSR news-specific public concern measure, I conduct an event study. The regression model is as follows. For CSR news from a firm i on a CSR activity dealing with social issue s, issued in time t,

$$\begin{aligned} CAR[-1,+1]_{i,s,t} =& \beta_0 + \beta_1 CSR\_Concern(M)_{i,s,t} + \mu_s + \epsilon_{i,s,t}, \\ & M \in \{3,6\} \end{aligned} \tag{9}$$

$CAR[-1,+1]$ is the Carhart (1997) four-factor adjusted cumulative abnormal returns summed over a three-day window including the day before the event, the event day, and the day after the event. In this analysis, I include issue fixed effects (μ_s) to hone in on the temporal variation of $CSR\_Concern(M)$. In other words, any differences in public concern metrics and market reactions attributable to the four issues are accounted for by controlling for the average differences. The positive coefficient of $CSR\_Concern(M)$ will indicate that stock market reaction is more favorable to corporate actions on a social issue in times when the social issue is more concerning.

Before running the analysis, it is worth noting that some CSR news releases contain information on multiple CSR activities. Moreover, there are cases where there is more than one CSR news release published on the same day from a company. This brings about two cases I need to handle separately. First, multiple CSR activities are reported in multiple news releases on the same day from a firm, but they all address the same type of social issue. In that case, I treat the news releases as duplicates and leave only one CSR activity news. Second, there are multiple CSR news releases but they address different social issues. To deal with such cases I build two samples. In the first sample, I retain one CSR activity for each distinct social issue, potentially counting multiple CSR activities in a day, as they are assigned different values for $CSR_Concern(M)$. However, the process leaves one CSR activity for an issue. In the second sample, I only keep a CSR activity that has the highest $CSR\_Concern(M)$ value, assuming the news is more important to investors. The results reported in the paper are based on the first sample. In analyses not reported in the paper, I conduct the same regression analyses using the second sample and find results that are both economically and statistically similar to those reported.

Table 4 presents the findings. When running regressions without any control variables or fixed effects aside from issue-fixed effects, both $CSR\_Concern(3)$ and $CSR\_Concern(6)$ show positive coefficients that are statistically significant at the 1% significance level. This positive relationship suggests that CSR announcements receive a more favorable market reaction when firms address social issues that are prominent in public discourse. Even when introducing either industry or firm fixed effects into the model, the coefficients tied to public concern metrics remain significant at the 5% significance level.

Specifically, in column (6), where issue- and firm-specific fixed effects are included, a one standard deviation increase in $CSR\_Concern(3)$ leads to a 0.061 percentage point increase in market reactions. In column (12), a one standard deviation increase in $CSR\_Concern(6)$ results in a 0.058 percentage point higher market response.[22] Taking into account that the average market reaction on a CSR activity is 0%, a one standard deviation rise in the three-month concern metric results in an average news return of 0.061%.

However, interpreting the economic significance of these coefficients is complex because the metrics for public concern are only approximations of the actual level of public concern. The standard deviations of the public concern metrics I use to assess the impact on market reactions may not be the same as the standard deviation of the true public concern. Therefore, without knowing the precise relationship between the true level of public concern and its proxy variable, my interpretation of the results is limited to the observed linear relationship between the metric and market reactions. Nevertheless, provided there is a linear relationship between the true and approximated measures, which is likely, we can assert that the level of public concern has a meaningful positive linear relationship with the market reactions to CSR activities.

Overall, the results in **Table 4** support the hypothesis. When companies address

[22]The standard deviation of $CSR\_Concern(3)$ de-meaned by issue is 0.153. The standard deviation of $CSR\_Concern(6)$ de-meaned by issue is 0.145.

social issues that align with heightened public concern, they elicit more favorable market reactions. As the level of public concern is linked to the demand for CSR, the findings suggest that there are times and contexts in which CSR can enhance firm value through its positive influence on the externalities component of incremental cash flows.

3.2.2 The Methods of CSR Activities and Market Reactions

In this subsection, I explore whether the market's response varies based on the approaches firms adopt to address environmental and societal challenges. If the value of CSR activities to a firm is a function of the public's demand for CSR, these activities must address those needs in a manner that aligns with the public's expectations of how corporations should tackle these issues to increase firm value.

Many environmental issues, such as pollution, waste, and emissions, stem directly from business operations.[23] Moreover the magnitude of the environmental challenges, from climate change to plastic waste, requires collaborative efforts from all sectors of society, including corporations. For inclusion and diversity issues, companies have a direct impact on the workforce as they employ a significant portion of the population. Their policies, practices, and cultures directly affect millions of workers. In addition, many businesses engage with local communities through their services and products. Environmental and inclusion challenges, therefore, may necessitate deeper engagement than mere financial donations to third-party organizations.

On the contrary, poverty and education might benefit from corporate philanthropy. These issues can see immediate benefits from financial injections. For instance, donations can provide school supplies, assist teachers, or directly provide resources to impoverished communities. Furthermore, it is challenging to directly link issues of poverty and education to the actions of corporations. This lack of clarity creates a gap in understanding the extent of a corporation's direct engagement in poverty and education issues, positioning corporate giving as one of the most feasible solutions. However, corporate donations to third-party organizations might be more easily replicated by individuals unless they encounter significant transaction costs in transferring money to the beneficiaries they aim to assist (Bénabou and Tirole (2010)).

Given this context, I propose a testable hypothesis: investors may value a com-

[23] According to a report by BBC, the industry is responsible for producing one-third of global waste (Miller, N. (2021) BBC Future). Moreover, based on data from the EPA, in 2021, industries were responsible for 23% of greenhouse gas emissions in the U.S. Meanwhile, transportation contributed to 28%, and electricity generation made up 25% of the emissions. Link to the report.

pany's monetary donations less when addressing environmental and inclusion challenges, as these may not be the most effective methods to meet public demands for resolving such issues. For poverty and education, the value of corporate giving could be either positive or negative. I introduce an indicator variable, denoted as *Giving*, which takes a value of one if the mode of a CSR activity involves corporate giving only and zero otherwise. Also, an indicator variable $I(Env|Inc)$ is created to capture issues that require more than financial contributions. Specifically, the variable takes one if a CSR activity addresses environmental or inclusion-related issues, and zero otherwise. Under the hypothesis, the interaction of *Giving* and $I(Env|Inc)$ must have a negative coefficient.

Table 5 presents the results. I incorporate year-fixed effects in all regressions to control for the influence of varying public concerns over social issues. I drop issue fixed effects to allow for the inclusion of $I(Env|Inc)$. In some regressions, I include industry or firm fixed effects.

The coefficient of *Giving* is not statistically significant on its own (columns (1) and (6)), showing corporate donations do not increase firm value on average. Likewise, the coefficient of $I(Env|Inc)$ is not statistically significant alone in columns (2) and (3), meaning addressing specific social issues itself is not a value-enhancing driver. When the interaction term is included, *Giving* exhibits positive coefficients in all regression specifications, though with limited statistical significance. For raw returns, the coefficient estimates are positive and reach statistical significance at the 10 percent level. However, for cumulative abnormal returns, the estimates do not achieve statistical significance. Nonetheless, these findings suggest there are instances where corporate giving is perceived as valuable in the areas of education and philanthropy, likely driven by transaction costs associated with individual giving (Bénabou and Tirole (2010)).

Focusing on the interaction term, the coefficients for this term are consistently negative and statistically significant across all regression specifications. Notably, the sum of the coefficients of the interaction term and *Giving* is negative. This indicates that corporate giving generally draws negative market reactions when it comes to environmental and inclusivity concerns. As detailed in column (10), when companies tackle these issues with strategies beyond mere financial donations, market reactions to CSR activities for those issues are more favorable by 0.3 percentage points compared to corporate giving.

The findings align with the hypothesis that factors related to public demand for CSR influence the market value of CSR activities. If a CSR activity is intended to respond to public demand for corporate action on certain issues, it should be executed in a manner that effectively addresses those issues and in a way that cannot be easily replicated by individuals. The findings also carry important implications concerning the issue of ‘cheap talk’ or ‘virtue signaling,’ which will be further addressed in the discussion section.

3.2.3 Partial Effects of Public Concern and Method of Addressing Social Issues

I further investigate whether the two variables related to the public’s needs for CSR retain their impact on the market’s reaction when considered simultaneously. For this purpose, I conduct regressions that only include firm fixed effects, enabling the inclusion of both variables.

Table 6 reports the results. The results suggest that these variables maintain their statistical as well as their economic significance. This suggests that when valuing CSR efforts, investors give considerable weight to both the level of public concern tied to a social issue and the method used to address it.

3.2.4 Additional Drivers of CSR Value

I delve deeper into the stock market valuation of CSR activities, particularly alongside other CSR-related variables. The goal is to determine if factors associated with public perception of CSR continue to influence the market’s valuation even when other variables are considered. Additionally, I explore the impact of those variables on the market valuation of CSR initiatives.

3.2.4.1. Financial Performance and Market Reactions to CSR Activities
In this subsection, I investigate the market’s reaction to CSR activities, taking into account the firm’s profitability. According to Friedman (1970), the primary responsibility of a business, specifically its managers, is to maximize profits for its shareholders. Embarking on CSR initiatives often demands significant corporate resources, from financial investments to managerial capacity. This diversion has the potential to adversely affect a firm’s financial outcomes. Should the incremental cash flows from CSR activities be solely tied to their related opportunity costs, then factoring in profitability metrics could eclipse factors associated with the public demand for CSR. This is due to the diminishing impact of CSR opportunity costs with increasing past profits and accumulated financial reserves.

For proxies of a firm’s profitability, I use the average return on assets over the past two years and EPS surprises on the most recent EPS announcement day within the last 180 days. EPS surprises are determined by deducting the median forecast from the actual quarterly EPS. This median prediction is derived from analyst forecasts made 2 to 15 days prior to the earnings report when available; otherwise, it is based on forecasts given 16 to 30 days before the disclosure. This measure is then adjusted according to the company's share price, using data closest to the EPS announcement date, specifically from five days to three days before the event.

Table 7 reports the results. First, the most recent financial news is a strong predictor of the market reaction in terms of statistical significance. The magnitude of the coefficients is also large. A one standard deviation increase in EPS surprises (0.011) leads to a 0.15 percentage point higher three-day cumulative return around the news on a CSR activity on average, controlling for issue- and firm-fixed effects. The same change in EPS surprises translates into 0.16 percentage points higher cumulative abnormal returns on average. Exhibiting lower statistical significance, an increase of one standard deviation in ROAs over the past two years corresponds to an enhancement of 0.14 percentage points in market reactions, measured as cumulative abnormal returns. The finding shows that investors factor in companies' financial performance since CSR requires companies' financial resources, incurring opportunity costs. This indicates that the financial slack of firms, or their capacity to execute CSR initiatives with minimal opportunity costs, constitutes a critical determinant for investor consideration.

Turning to the effect of the factors related to the public demand for CSR, the coefficients for the variables gauging public concern remain significant, with only minor variations in their magnitudes. This suggests that a firm's profitability does not completely overshadow the effect of public concern on the market reaction to CSR programs. Also, the results show the method of CSR initiatives remains a crucial factor in investor evaluations, drawing more positive market reactions to corporate actions other than donations for environmental and diversity-related issues. Overall, while financial performance is a crucial factor that determines the value of CSR activities, the potential of CSR to increase firm value still matters.

3.2.4.2. Governance Structure and Market Reactions to CSR Activities

The CSR literature has noted the relationship between CSR and corporate governance. One of the prevalent views is that CSR activities could reflect managerial agency problems. For example, Masulis and Reza (2014) find that higher corporate donations lead to lower shareholder valuation of a firm's cash holdings, and suggests that such donations may align with CEO interests, potentially misusing corporate resources and diminishing firm value. Also, Di Giuli and Kostovetsky (2014) show that companies with higher CSR scores often reflect the political preferences of their executives and directors. Additionally, these CSR ratings correlate with future reductions in stock returns and a decrease in the firm's ROA.

Conversely, some perspectives suggest that governance might influence both the adoption and valuation of CSR initiatives. Cheng, Hong, and Shue (2013) find an increase in monitoring leads to lower ESG scores. Ferrell, Liang, and Renneboog (2016) show that well-governed firms with fewer agency issues are more engaged in CSR and implement it in a way that these efforts are less affected by agency issues, which positively relates to value and mitigates the negative effects of managerial entrenchment on value.

If the value and initiation of CSR activities can be mainly attributed to agency costs, then adding variables that capture corporate governance might overshadow the effects of factors introduced in the paper. I construct several variables tied to the board structure. $chair\_CEO$ is a binary variable that equals one if the CEO also serves as the board chairman, and zero otherwise. It gauges the degree to which a CEO's personal preferences influence CSR participation and its subsequent impact on company value. $log(\#director)$ represents the logarithm of the board's size plus one. The motivation behind this variable is rooted in the finding that having a larger board can erode company value, arising from inefficient communication and decision-making challenges typically associated with it (Guest (2009)). Lastly, $frac\_ind$ measures the fraction of the board comprised of independent directors. It signifies the positive value these independent directors offer to shareholders, as documented by Nguyen and Nielsen (2010).

Table 8 presents the results. In line with the notion that CSR initiatives, when heavily influenced by CEO preferences, may not be positively received, the coefficients of $chair\_CEO$ consistently emerge as negative and statistically significant across all regression models. Regardless of the specifics of model design, when the CEO doubles as the board chairman in firms rolling out CSR endeavors, the market reaction dwindles by 0.2 percentage points. As expected, $log(\#director)$ carries negative coefficients, yet they do not attain statistical significance. Similarly, $frac\_ind$ also fails to exhibit significant coefficients although showing positive coefficients.

After accounting for board structure attributes, the coefficients representing public concern and the effectiveness of CSR initiatives retain their statistical significance. The magnitude of these coefficients remains similar to what we observe in the absence of board characteristics. This suggests that the degree of public concern and the approach firms take to act on social issues remain influential determinants independent of board composition in shaping the market's valuation of CSR initiatives.

3.3 Corporate Decisions on CSR Initiatives

The preceding section highlights the possibility of enhancing firm value by addressing environmental and societal challenges. In this section, the analysis focuses on whether companies consider factors that could enhance the value of CSR initiatives in their decision-making.

3.3.1 Public Concern and CSR Activities

In this subsection, I investigate the likelihood of corporations increasing their CSR efforts in accordance with the level of public concern on specific issues, as such endeavors might positively affect their overall value.

A visual inspection of **Figure 1** supports the hypothesis. In Panel B of Figure 1, I juxtapose the public concern index over diversity-related issues with the count of CSR initiatives addressing these concerns. Notably, both the public concern metric and CSR actions saw a marked rise in the month following the George Floyd protest in May 2020. This event notably energized the Black Lives Matter movement and influenced various sectors of society to reevaluate issues of race and equality. The concurrent rise in the number of articles covering diversity issues in the op-ed sections and the reported CSR initiatives related to the issues in the period hints at how companies react to critical social issues.

To examine if companies strategically time their actions on environmental and social issues, I construct a dataset where the unit of observation is firm-month. I create an indicator variable representing CSR activities that address specific social issues($Issue$, $Issue \in \{Environment,\ Inclusion,\ Education,\ Poverty\}$) in a given month. Specifically, $I(CSR : Issue)_t$ is an indicator variable that is assigned a value of one if, during the month t, a firm reports a news release that introduces a CSR initiative related to a specific social issue ($Issue$).

The primary independent variable is the level of public concern regarding a specific social issue, quantified over various timeframes: the current month, the preceding three months, the preceding six months, or the preceding twelve months, with the current month included in each period. The coefficient of the variable indicates whether companies are responsive to issues that have recently garnered public attention.

The model specification is, for a firm i in time t,

$$\begin{aligned}
&I(CSR:Issue)_{i,t} = \beta_0 + \beta_1 Concern_{t-k,t,SI} + \gamma^{'} X_{i,t} + \epsilon_{i,t}, \\
&Issue \in \{Environment,\ Inclusion,\ Education,\ Poverty\} \\
&k \in 1, 3, 6, 12
\end{aligned} \tag{10}$$

Table 9 shows whether corporations choose to address issues in response to prevailing public concerns. Across all issues, there is a pronounced link between the deployment of pertinent CSR activities and the level of public concern. For environmental issues, a one standard deviation increase in public concern—averaged over the past twelve months (0.47) - increases the likelihood of launching CSR actions tied to the environment by 0.188 percentage points. This represents a 15.67% increased probability relative to the unconditional likelihood of introducing environmental CSR initiatives. For inclusion issues, a one standard deviation increase in public concern over the past year (0.39) raises the probability of implementing inclusion-focused CSR by 0.23 percentage points. This translates to a 39.00% heightened likelihood relative to the unconditional probability of commencing inclusion CSR activities. when it comes to education, a one standard deviation rise in public concern over the past year (0.4) increases the odds of undertaking education-related CSR by 0.16 percentage points. This equates to a 12.3% enhanced probability compared to the unconditional chance of rolling out educational initiatives. Lastly, regarding poverty issues, a one standard deviation growth in public concern over the past twelve months (0.42) increases the chance of launching poverty-associated CSR by 0.21 percentage points. This corresponds to an 11.7% higher likelihood relative to the baseline probability of initiating poverty-focused initiatives.

The results offer insight into corporations' strategic reactions to dominant societal issues, aligning closely with the notion that firms are inclined to address social matters prominent in public discourse. Although not directly examined in this section, the analysis suggests that companies may anticipate improving their value through active engagement with such issues. Alternatively, it may indicate that managerial motivations align with increasing CSR activities in response to public concerns. Irrespective of the underlying motives, the analysis reveals a robust correlation between public concern regarding an issue and firms' decisions to engage with that issue.

3.3.2 Corporate Decisions on the Methods of Addressing Issues

In this section, I explore whether firms choose methods to address critical issues in a manner that positively impacts their stock prices. Specifically, I examine whether companies adopt CSR initiatives that extend beyond mere financial contributions when addressing environmental and inclusion-related challenges. Additionally, I investigate whether the choice of approach is influenced by the level of public concern. As issues become increasingly pressing, companies may opt for more effective methods to address them.

I move back to the sample of press releases on CSR activities regarding the four issues. I run regressions where the dependent variable is the indicator for corporate giving (*Giving*). Independent variables are the level of public concern, a variable singling out environmental and inclusion issues, and an interaction term of them.

Table 10 reports the results. First, when $I(Env|Inc)$ is the sole predictor in a regression model, it receives a coefficient that is negative and statistically significant at the 1% significance level. Columns (1) and (2) show that companies are 45% less likely to opt for initiatives that only involve financial contributions when addressing environmental or inclusion issues.

Second, I examine whether the level of public concern influences the decisions companies make regarding their methods of CSR initiatives. In regressions where $I(Env|Inc)$, the level of public concern, and the interaction term are all included, I find that companies are more inclined towards monetary donations when addressing poverty and education issues particularly as these issues receive more public attention. This is shown by the positive and statistically significant coefficients of $CSR\_Concern(3)$ and $CSR\_Concern(6)$. Regarding environmental and inclusion-related issues, there is a noticeable preference for CSR initiatives that extend beyond financial contributions to charities or NGOs, especially as these issues attract public interest. This is evidenced by the negative sum of the coefficients of the interaction term and the level of public concern (e.g., $I(Env|Inc) \times CSR\_Concern(3)$ and $CSR\_Concern(3)$).

The findings reveal a clear pattern in corporate approaches to CSR, influenced by the type of social issue and the level of associated public concern. Firms tend to choose financial contributions when addressing issues such as education and poverty, particularly when these matters become more prominent in public discourse. In contrast, when dealing with environmental and diversity-related challenges, companies lean towards a more involved approach. This strategic selection of CSR methods may assist firms in enhancing their value.

3.4 Discussion

In this section, I discuss the average market reaction to CSR initiatives, which is measured by the average CAR around press releases, that hovers around zero. I propose four explanations for the observation. These explanations are based on the evidence detailed in the paper and the Online Appendix.

First, CSR activities that might serve as virtue-signalling may not enhance firm value. This can be inferred from the findings related to corporate giving. In the sample, donations takes up 52% of all CSR activities (**Table 3**). However, on average, corporate giving does not generate additional value for firms (**Table 5**). Importantly, corporate donations linked to environmental or inclusion-related concerns result in a notable decline in stock prices. Conversely, market responses to financial contributions focused on poverty and education are mildly positive or lukewarm. This raises questions about the capacity of corporate giving to enhance firm value and the reasons why the majority of CSR activities are centered around it.

Corporate giving can be less costly compared to initiatives like operational overhauls, project implementations, or adopting new cultural policies. This is due to its ease of implementation, potential tax benefits, minimal business disruption, and flexibility in terms of donation amount and frequency. For companies seeking a cost-effective way to demonstrate social responsibility, donating to a third-party organization could be a tempting option.

However, other forms of CSR initiatives can offer more sincere gestures to provide solutions to the issues companies aim to address. If public demand for CSR plays a crucial role in determining the value of CSR activities to a firm, merely sending a superficial signal of social responsibility through financial contributions may not sufficiently influence consumers, regulatory bodies, and investors alike. Ultimately, discerning investors may be able to distinguish between "virtue signaling" through donations and genuinely impactful CSR practices, reflecting this understanding in stock prices. This distinction could explain the zero average market reactions to both corporate giving and CSR initiatives.

Second, although profitability is crucial in driving market reactions to CSR activities, it seems to have little bearing on a company's engagement in CSR initiatives. In **Table 7**, two-year average ROAs affect market reactions positively at the 10 percent significance level. However, in further investigations detailed in the Online Appendix (**Table OA.2**), it is found that the same variable does not affect the firms' decision to conduct CSR activities. This disconnect suggests that firms may not fully appreciate how profitability should guide CSR objectives. If there were more cases where firms had either participated or abstained from participating in the CSR movement based on their profitability, it could have potentially resulted in a positive average market reaction to CSR activities.

Third, the influence of the CEO within the board, a factor that negatively predicts the market value of CSR (**Table 8**), shows a positive correlation with philanthropic CSR, including activities focused on education and poverty. In a regression presented in the Online Appendix (**Table OA.2**), the inclusion of a dummy variable representing the CEO-Chairman dual role yields a positive coefficient (0.018, t-statistic of 1.848) concerning involvement in CSR initiatives associated with education and poverty-related issues. This coefficient suggests a 11.58% increase in the likelihood of engaging in such activities compared to the unconditional probability.[24] As evidenced by prior research, there exists a concern regarding CEOs participating in these activities for personal gain (Bénabou and Tirole (2010), Cheng, Hong, and Shue (2013), Masulis and Reza (2014), Di Giuli and Kostovetsky (2014)). The observation that powerful CEOs tend to engage in areas where corporate giving constitutes the majority of CSR activities suggests that these initiatives, presumably influenced by CEOs, may signify decisions that stray from the fundamental goal of maximizing firm value.

Finally, there could be peer effects in the adoption of CSR activities. In the Online Appendix, I show that two types of CSR initiatives, environmental and poverty-related CSR activities, demonstrate a statistically significant presence of peer effects. This suggests that peer participation in those areas encourages a firm to adopt similar CSR practices. For general CSR activities, which include activities related to all types of issues, a one standard deviation rise in the proportion of peers engaging in CSR initiatives leads to a 0.047 increase in the likelihood of the firm itself undertaking a CSR initiative in the subsequent year. This represents a 16% increase from the unconditional probability of engaging in CSR activities in any given year (29.36%). The findings suggest that some companies may undertake CSR initiatives without a thoroughly planned strategy aimed at enhancing firm value, but simply because their peers are engaging in similar activities. This observation could potentially explain why we observe zero market reactions on average.

In summary, while public demand for CSR is a significant factor positively impacting the value of CSR activities, there are other motivations behind firms' engagement in CSR. Some of these motivations may inadvertently diminish firm value, contributing to why we observe neutral market reactions to CSR activities on average.

[24] In the sample, the unconditional probability of a firm engaging in CSR activities related to education and poverty in a given year is 15.55%.

3.5 Conclusion

This study pioneers a systematic analysis of corporate social responsibility (CSR) endeavors. The study identifies factors that influence the stock market valuations of these CSR activities. Like any other business project within a firm, the value of CSR activities must be determined by considering their incremental cash flows. These cash flows should reflect opportunity costs, as well as the negative and positive externalities on the firm's ongoing operations. For CSR activities to be considered valuable, it should carry the positive externalities affecting operations.

The focus of this study is on a factor that influences the incremental cash flows of a CSR activity through positive externalities: public demand for CSR. I posit that public demand for CSR intensifies when society grapples with heightened concerns over social and environmental issues, also when these challenges overwhelm individual capabilities. In such situations, the public provides incentives to firms to actively engage in social issues. This active engagement can, in turn, increase the value of the participating firms when the public need for CSR is high.

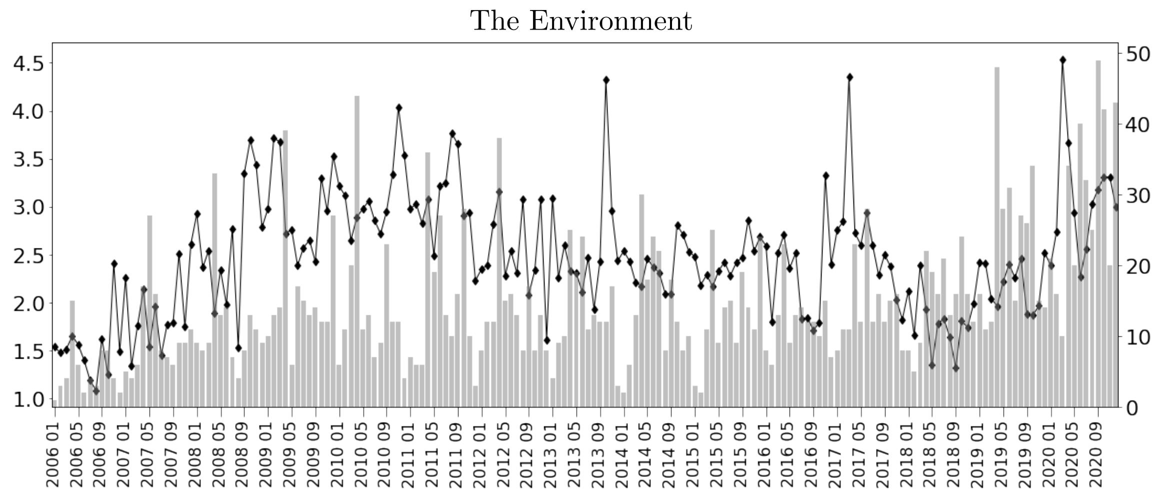
The study examines market reactions to CSR activities, measured by Carhart's four-factor adjusted cumulative abnormal returns. The findings reveal that, on average, market reactions to CSR activities are neutral. However, the level of public concern, as indicated by the frequency of articles discussing specific social issues, influences market reactions to corresponding CSR activities. Moreover, CSR strategies that address societal challenges more effectively than individuals can elicit more favorable market responses. These findings remain significant even after accounting for factors that traditionally explain the adoption and value proposition of CSR initiatives, such as profitability and corporate governance.

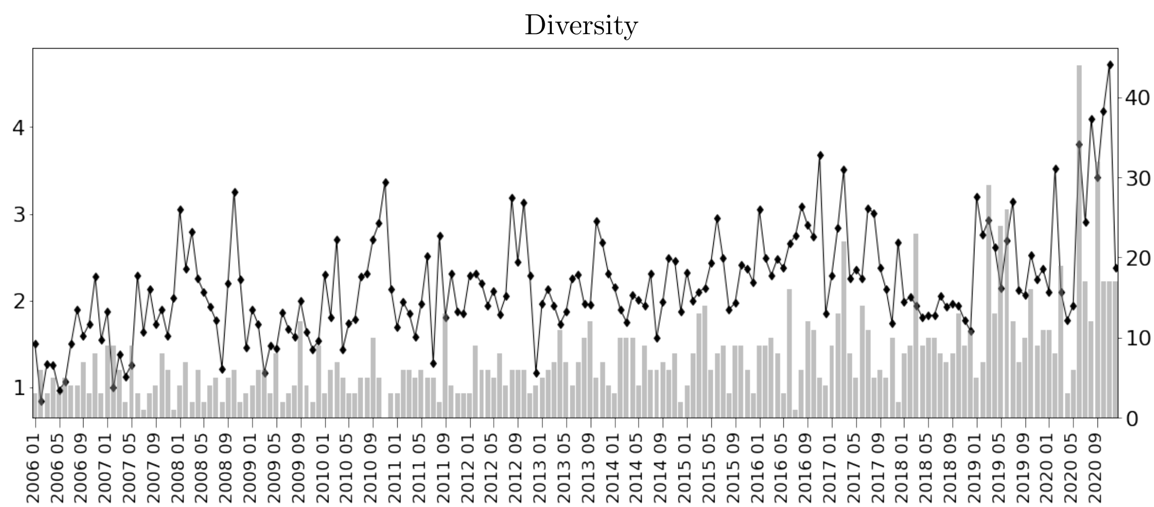
Delving deeper, the study explores the strategic CSR conduct of companies aiming to enhance their value. Firms tend to escalate their CSR activities in response to urgent social issues and strategically determine their optimal methods to these challenges. The study also discusses why market reactions to CSR activities are, on average, neutral despite such strategic CSR implementation.

Overall, this paper offers insights into how CSR initiatives can enhance firm value. It suggests that CSR activities, when implemented in ways that garner widespread public support, can positively influence stock prices.

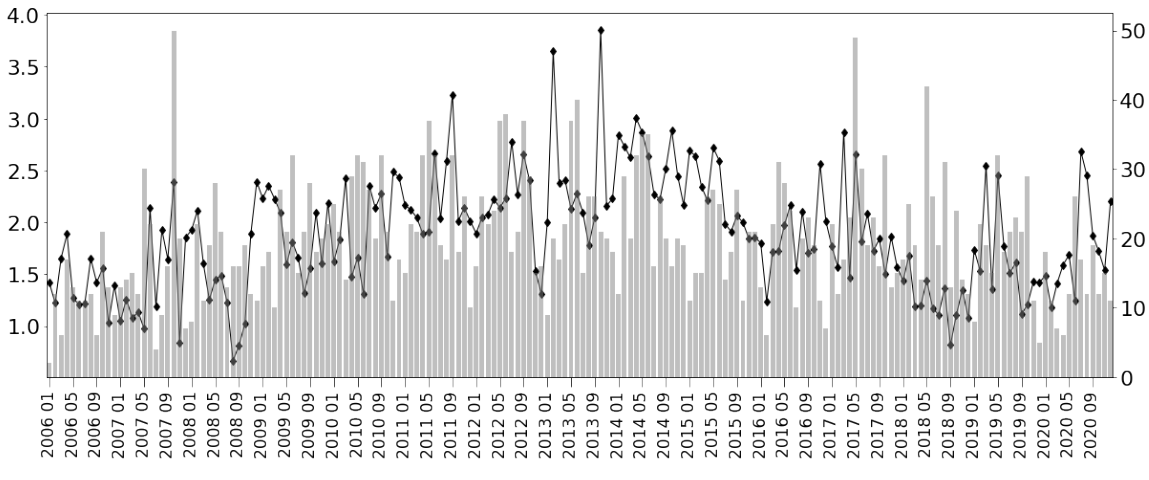
Figure 1. Public Concern Indices and the Number of CSR Activities

The four figures below plot the monthly indices of public concern about four social and environmental issues: the environment, diversity, education, and poverty in the US. The indices represent the source-normalized number of articles covering each issue in the op-ed sections of The New York Times, The Wall Street Journal, and USA Today. The bar represents the number of news releases on CSR activities related to a specific issue, as specified in the title of each panel. The left y-axis represents the monthly index of public concern over an issue, while the right y-axis shows the number of news releases on CSR activities addressing that issue. The x-axis represents time, with each point indicating a specific month and year.





Education



Poverty in the U.S.

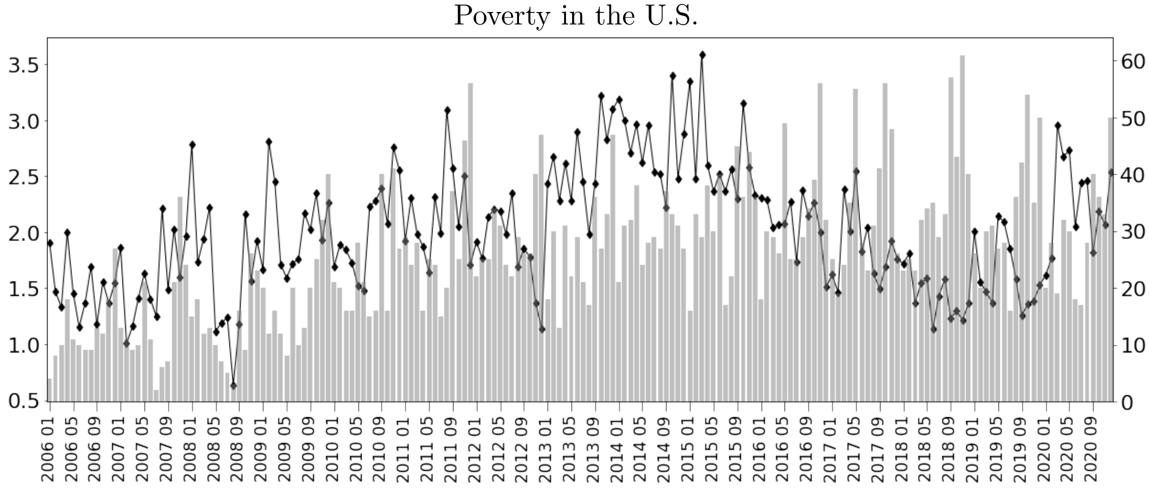


Table 1. Descriptive Statistics of CSR Activities

This table displays the annual frequency of CSR activities reported by firms and the count of firms engaging in CSR activities. Additionally, it provides descriptive statistics on annual CSR activities per firm for those that reported at least one CSR activity, including the mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum values. The sample is compiled by analyzing press releases from 1,430 companies over the period from 2006 to 2020.

		CSR Frequency		Annual Count of CSR Activities by Firms						
Year	Firms	Activities	Firms	Mean	Std	Min	25th	Median	75th	Max
2006	1112	758	190	4.10	9.05	1	1	2	3	86
2007	1107	1068	220	4.94	14.68	1	1	2	4	148
2008	1100	1022	244	4.25	9.64	1	1	2	4	102
2009	1106	1216	270	4.51	10.75	1	1	2	4	138
2010	1129	1529	314	4.79	12.17	1	1	2	4	174
2011	1141	1702	354	4.79	10.84	1	1	2	4	156
2012	1188	1752	382	4.59	8.40	1	1	2	5	101
2013	1199	1588	374	4.14	6.65	1	1	2	4	63
2014	1195	1587	384	4.18	8.26	1	1	2	4	106
2015	1164	1588	400	3.97	6.12	1	1	2	4	58
2016	1131	1681	415	4.00	6.33	1	1	2	4	61
2017	1100	1925	457	4.30	8.64	1	1	2	5	145
2018	1058	1907	463	4.18	8.00	1	1	2	4	114
2019	1025	2048	525	3.83	7.26	1	1	2	4	122
2020	989	2327	586	3.81	4.98	1	1	2	4	53

Table 2. Market Reactions to CSR Activities

The table reports descriptive statistics for cumulative returns and cumulative abnormal returns around news releases on CSR activities, segmented into overall and category-specific (Environment, Inclusion, Education, Philanthropy) analyses. N represents the number of observations, followed by the mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum. Cumulative returns are computed over a three-day window surrounding the event day, defined as the day a news release on CSR activities is published. Cumulative abnormal returns are adjusted for risk using the Carhart (1997) four-factor model, calculated over three days surrounding the event day. Returns are winsorized at the 1st and 99th percentiles.

Variable	N	Mean	Std	Min	25th	Median	75th	Max
Cumulative Returns around news releases on CSR activities ($Return[-1,+1]$)								
All	21924	0.002	0.031	-0.095	-0.013	0.002	0.016	0.108
Environment	2984	0.002	0.032	-0.095	-0.014	0.002	0.017	0.108
Inclusion	2056	0.002	0.030	-0.095	-0.012	0.003	0.017	0.108
Education	3243	0.001	0.028	-0.095	-0.013	0.002	0.015	0.108
Philanthropy	3747	0.002	0.031	-0.095	-0.013	0.002	0.017	0.108
Cumulative Abnormal Returns around news releases on CSR activities ($CAR[-1,+1]$)								
All	21924	0.000	0.024	-0.079	-0.012	0.000	0.011	0.085
Environment	2984	0.000	0.025	-0.079	-0.012	-0.001	0.011	0.085
Inclusion	2056	0.000	0.023	-0.079	-0.011	-0.001	0.012	0.085
Education	3243	0.000	0.022	-0.079	-0.011	0.000	0.011	0.085
Philanthropy	3747	0.000	0.025	-0.079	-0.012	0.000	0.012	0.08

Table 3. Summary Statistics

This table presents the number of observations (N), along with the mean, standard deviation, minimum, 25th percentile, median, 75th percentile, and maximum for the variables used in the paper. $Return[-1,+1]$ and $CAR[-1,+1]$ represent the cumulative returns and cumulative abnormal returns associated with news releases on CSR activities, respectively. The variable $CSR\_Concern(M)$, where M is either 3 or 6, indicates the level of public concern for an issue that a CSR activity targets, measured over the preceding M months. *Giving* is an indicator variable set to one when a CSR activity involves corporate giving only, and zero otherwise. $I(Env|Inc)$ is an indicator variable that is assigned a value of one if a CSR activity is related to environmental or diversity issues, and zero otherwise. Avg_roa_2y is the average return on assets (ROA) in the past two years. $EPSsurprise$ is EPS surprise defined as the difference between actual EPS and the median analysts' forecast. $chair\_CEO$ is an indicator variable that takes one when a CEO is the chairman of the board, and zero otherwise. $log(\#director)$ is the logarithm of one plus the number of directors on a company's board. $frac\_ind$ is the proportion of independent directors on the board. $I(CSR : Issue)$ takes a value of one if a firm reports a CSR activity addressing an issue ($Issue \in \{Environment,\ Inclusion,\ Education,\ Poverty\}$) in the month, and zero otherwise. $Concern_{t-k,t,Issue}$ is the monthly index of public concern related to an issue ($Issue$), measured over the past k months.

Variable	N	Mean	Std	Min	25th	Median	75th	Max
New release observations								
$Return[-1,+1]$	10100	0.002	0.031	-0.095	-0.013	0.002	0.016	0.108
$CAR[-1,+1]$	10100	0.000	0.024	-0.079	-0.012	0.000	0.012	0.085
$CSR\_Concern(3)$	10149	1.138	0.173	0.606	1.021	1.146	1.258	1.631
$CSR\_Concern(6)$	10149	1.136	0.161	0.604	1.023	1.142	1.250	1.580
$Giving$	10149	0.521	0.500	0.000	0.000	1.000	1.000	1.000
$I(Env\|Inc)$	10149	0.357	0.479	0.000	0.000	0.000	1.000	1.000
$I(Env\|Inc) * Giving$	10149	0.074	0.262	0.000	0.000	0.000	0.000	1.000
Avg_roa_2y	11235	0.152	0.076	-0.457	0.110	0.147	0.183	0.470
$EPSsurpise$	9492	0.001	0.010	-0.070	0.000	0.000	0.002	0.061
$chair\_CEO$	9667	0.475	0.499	0.000	0.000	0.000	1.000	1.000
$log(\#director)$	9667	1.801	0.610	0.000	1.609	1.946	2.197	3.045
$frac\_ind$	9667	0.775	0.179	0.000	0.700	0.800	0.889	1.000

Table 3-Continued

Variable	N	Mean	Std	Min	25th	Median	75th	Max
Firm-month observations								
$I(CSR: Environment)$	199296	0.010	0.099	0.000	0.000	0.000	0.000	1.000
$I(CSR: Inclusion)$	199296	0.008	0.091	0.000	0.000	0.000	0.000	1.000
$I(CSR: Education)$	199296	0.011	0.106	0.000	0.000	0.000	0.000	1.000
$I(CSR: Poverty)$	199296	0.015	0.120	0.000	0.000	0.000	0.000	1.000
$Concern_{t-1,t,Environment}$	199296	2.469	0.632	1.084	2.039	2.426	2.862	4.537
$Concern_{t-3,t,Environment}$	199296	2.460	0.542	1.223	2.117	2.397	2.839	3.714
$Concern_{t-6,t,Environment}$	199296	2.449	0.495	1.350	2.171	2.449	2.905	3.381
$Concern_{t-12,t,Environment}$	199296	2.424	0.472	1.359	2.157	2.468	2.805	3.265
$Concern_{t-1,t,Inclusion}$	199296	2.164	0.603	0.839	1.805	2.092	2.436	4.725
$Concern_{t-3,t,Inclusion}$	199296	2.154	0.487	1.093	1.862	2.092	2.482	4.108
$Concern_{t-6,t,Inclusion}$	199296	2.134	0.421	1.147	1.904	2.113	2.415	3.854
$Concern_{t-12,t,Inclusion}$	199296	2.100	0.382	1.084	1.932	2.145	2.285	3.078
$Concern_{t-1,t,Education}$	199296	1.896	0.555	0.667	1.487	1.886	2.231	3.858
$Concern_{t-3,t,Education}$	199296	1.889	0.462	0.834	1.540	1.911	2.224	2.838
$Concern_{t-3,t,Education}$	199296	1.882	0.429	0.829	1.522	1.875	2.212	2.786
$Concern_{t-12,t,Education}$	198427	1.876	0.401	1.168	1.529	1.856	2.220	2.665
$Concern_{t-1,t,Poverty}$	199296	2.023	0.552	0.636	1.586	1.998	2.381	3.590
$Concern_{t-3,t,Poverty}$	199296	2.016	0.479	1.016	1.636	1.985	2.394	3.136
$Concern_{t-6,t,Poverty}$	199296	2.008	0.448	1.253	1.662	1.992	2.299	3.029
$Concern_{t-12,t,Poverty}$	198427	1.992	0.424	1.255	1.638	1.964	2.204	2.839

Table 4. Public Concern and Market Reactions to CSR Activities

This table shows the impact of public concern about social or environmental issues on market reactions to CSR activities, based on a sample of press releases addressing four issues: environment, inclusion, education, and poverty in the U.S. The dependent variable $Return[-1,+1]$ represents cumulative returns over a three-day event window, centered around the release of news about a CSR activity. The dependent variable $CAR[-1,+1]$ represents Carhart (1997) four-factor adjusted cumulative abnormal returns over a three-day event window. $CSR\_Concern(M)$, with $M \in \{3, 6\}$, represents the level of public concern for an issue being addressed by a CSR activity over the past M months. All regressions include fixed effects for the four issues. 'Firm FE' and 'Industry FE' represent firm and industry fixed effects, respectively. The industry is defined as the first three digits of the SIC. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	$Return[-1,+1]$			$CAR[-1,+1]$		
	(1)	(2)	(3)	(4)	(5)	(6)
$CSR\_Concern(3)$	0.009***	0.009***	0.009***	0.004***	0.004***	0.004**
	[4.53]	[4.40]	[4.00]	[2.76]	[2.67]	[2.25]
Issue FE	✓	✓	✓	✓	✓	✓
Industry FE		✓			✓	
Firm FE			✓			✓
Adjusted R^2	0.002	0.015	0.034	0.0004	0.008	0.029
Observations	10100	10051	9935	10100	10051	9935
	$Return[-1,+1]$			$CAR[-1,+1]$		
	(7)	(8)	(9)	(10)	(11)	(12)
$CSR\_Concern(6)$	0.010***	0.010***	0.009***	0.004***	0.004**	0.003**
	[4.73]	[4.44]	[4.02]	[2.62]	[2.40]	[2.04]
Issue FE	✓	✓	✓	✓	✓	✓
Industry FE		✓			✓	
Firm FE			✓			✓
Adjusted R^2	0.002	0.015	0.034	0.0003	0.007	0.029
Observations	10100	10051	9935	10100	10051	9935

Table 5. CSR Implementation Modes and Market Reactions

This table shows the effect of the methods to act on social and environmental issues on market reactions to CSR activities. The sample contains press releases on CSR activities that focus on four societal and environmental issues (environment, inclusion, education, and poverty in the U.S.). The dependent variable $Return[-1,+1]$ represents cumulative returns over a three-day event window, centered around the release of news about a CSR activity. The dependent variable $CAR[-1,+1]$ represents Carhart (1997) four-factor adjusted cumulative abnormal returns over a three-day event window. *Giving* is an indicator variable that assumes a value of one if a CSR activity is corporate giving, and zero otherwise. $I(Env|Inc)$ is an indicator variable that takes a value of one if a CSR activity pertains to either environmental concerns or issues related to diversity, and zero otherwise. 'Firm FE' and 'Industry FE' stand for firm and industry fixed effects, respectively. The industry is defined as the first three digits of the SIC. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Return[-1, +1]					CAR[-1,+1]				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$I(Env\|Inc) \times Giving$			-0.004**	-0.004***	-0.004**			-0.003**	-0.003**	-0.003**
			[-2.52]	[-2.77]	[-2.54]			[-2.22]	[-2.22]	[-2.07]
Giving	0		0.001	0.002**	0.002**	0		0.001	0.001	0.001
	[0.64]		[1.48]	[2.16]	[2.17]	[0.25]		[1.05]	[1.53]	[1.40]
$I(Env\|Inc)$		0	0.001	0.002*	0.001		0	0.001	0.001	0.001
		[-0.28]	[1.27]	[1.72]	[1.58]		[-0.67]	[1.01]	[1.21]	[0.91]
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE				✓					✓	
Firm FE	✓	✓			✓	✓	✓			✓
Adjusted R^2	0.037	0.034	0.005	0.018	0.037	0.03	0.027	0.001	0.008	0.03
Observations	9935	9703	10100	10051	9935	9935	9703	10100	10051	9935

Table 6. Partial Effects of Public Concern and Implementation Modes

This table displays the partial effects of the variable that gauges the level of public concern related to a social issue tackled by a CSR activity, as well as the method used to address the issue. The sample includes news releases on CSR activities that address four issues: the environment, inclusion, education, and poverty within the U.S. The dependent variable $Return[-1,+1]$ represents cumulative returns over a three-day event window, centered around the release of news about a CSR activity. The dependent variable $CAR[-1,+1]$ represents Carhart (1997) four-factor adjusted cumulative abnormal returns over a three-day event window. $CSR\_Concern(M)$, with $M \in \{3,6\}$, represents the level of public concern for an issue being addressed by a CSR activity over the past M months. *Giving* is an indicator variable that assumes a value of one if a CSR activity is corporate giving, and zero otherwise. $I(Env|Inc)$ is an indicator variable that takes a value of one if a CSR activity pertains to either environmental concerns or issues related to diversity, and zero otherwise. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	$Return[-1,+1]$		$CAR[-1,+1]$	
	(1)	(2)	(3)	(4)
$CSR\_Concern(3)$	0.009***		0.004**	
	[4.35]		[2.34]	
$CSR\_Concern(6)$		0.010***		0.003**
		[4.34]		[2.14]
$I(Env\|Inc) \times Giving$	-0.004**	-0.004**	-0.003**	-0.003**
	[-2.46]	[-2.45]	[-2.03]	[-2.04]
Giving	0.002**	0.002**	0.001	0.001
	[1.99]	[2.00]	[1.27]	[1.29]
$I(Env\|Inc)$	0	0	0	0
	[0.23]	[0.14]	[0.14]	[0.17]
Firm FE	✓	✓	✓	✓
Adjusted R^2	0.032	0.032	0.027	0.027
Observations	9703	9703	9703	9703

Table 7. Past Financial Performance and Market Reactions to CSR Activities

This table shows the effect of a firm's previous financial performance on market reactions to CSR activities. The sample includes news releases about CSR programs that address four issues: the environment, inclusion, education, and poverty within the U.S. The dependent variable $Return[-1,+1]$ represents cumulative returns over a three-day event window, centered around the release of news about a CSR activity. The dependent variable $CAR[-1,+1]$ represents Carhart (1997) four-factor adjusted cumulative abnormal returns over a three-day event window. Avg_roa_2y is the average return on assets (ROA) in the past two years. $EPSsurprise$ is EPS surprise defined as the difference between actual EPS and the median analysts' forecast. $CSR\_Concern(M)$, where M is either 3 or 6, indicates the level of public concern for an issue that a CSR activity targets, measured over the preceding M months. $Giving$ is an indicator variable set to one when a CSR activity involves corporate giving only, and zero otherwise. $I(Env|Inc)$ is an indicator variable that is assigned a value of one if a CSR program is related to environmental or diversity issues, and zero otherwise. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	$Return[-1,+1]$				$CAR[-1,+1]$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Avg_roa_2y	0.007	0.007	0.007	0.009	0.017*	0.017*	0.017*	0.018*
	[0.69]	[0.68]	[0.65]	[0.79]	[1.70]	[1.69]	[1.68]	[1.84]
$EPSsurprise$	0.137**	0.138**	0.138**	0.139**	0.144***	0.144***	0.144***	0.149***
	[2.26]	[2.25]	[2.21]	[2.21]	[3.80]	[3.80]	[3.79]	[3.86]
$CSR\_Concern(3)$	0.011***				0.004**			
	[4.40]				[2.25]			
$CSR\_Concern(6)$		0.012***				0.003*		
		[4.44]				[1.79]		
$I(Env\|Inc) \times Giving$			-0.004**	-0.004**			-0.003**	-0.003*
			[-2.53]	[-2.42]			[-2.05]	[-1.96]
$Giving$			0.002**	0.002**			0.001*	0.001
			[2.09]	[2.05]			[1.69]	[1.65]
$I(Env\|Inc)$			0.002**	0.002**			0.001	0.001
			[2.06]	[2.10]			[1.40]	[1.33]
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Issue FE	✓	✓			✓	✓		
Year FE				✓				✓
Adjusted R^2	0.031	0.031	0.028	0.033	0.025	0.025	0.025	0.026
Observations	8093	8093	8093	8093	8093	8093	8093	8093

Table 8. Governance Structure and Market Reactions to CSR Activities

This table shows the effect of a firm's governance structure on market reactions to CSR activities. The sample contains news releases on CSR programs addressing four issues (the environment, inclusion, education, and poverty in the U.S.). The dependent variable $Return[-1,+1]$ represents cumulative returns over a three-day event window, centered around the release of news about a CSR activity. The dependent variable $CAR[-1,+1]$ represents Carhart (1997) four-factor adjusted cumulative abnormal returns over a three-day event window. $chair\_CEO$ is an indicator variable that takes one when a CEO is the chairman of the board, and zero otherwise. $log(\#director)$ is the logarithm of one plus the number of directors on a company's board. $frac\_ind$ is the proportion of independent directors on the board. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Return[-1, +1]				CAR[-1,+1]			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$chair\_CEO$	-0.002*	-0.002*	-0.002**	-0.002**	-0.002***	-0.002***	-0.002***	-0.002***
	[-1.85]	[-1.86]	[-2.06]	[-2.06]	[-2.99]	[-2.99]	[-3.06]	[-3.06]
$log(\#director)$	-0.002**	-0.002**	-0.001	-0.001	-0.001	-0.001	0	0
	[-2.19]	[-2.26]	[-1.51]	[-1.51]	[-0.98]	[-1.00]	[-0.35]	[-0.35]
$frac\_ind$	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001
	[0.92]	[0.93]	[0.97]	[0.97]	[0.42]	[0.42]	[0.44]	[0.44]
$CSR\_Concern(3)$	0.010***				0.004***			
	[4.44]				[2.67]			
$CSR\_Concern(6)$		0.010***				0.004**		
		[4.53]				[2.46]		
$I(Env\|Inc) \times Giving$			-0.004***	-0.004***			-0.003**	-0.003**
			[-2.82]	[-2.82]			[-2.49]	[-2.49]
$Giving$			0.002**	0.002**			0.001	0.001
			[2.09]	[2.09]			[1.29]	[1.29]
$I(Env\|Inc)$			0.002	0.002			0.001	0.001
			[1.63]	[1.63]			[1.09]	[1.09]
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓
Type FE	✓	✓			✓	✓		
Year FE			✓	✓			✓	✓
Adjusted R^2	0.016	0.016	0.019	0.019	0.01	0.01	0.011	0.011
Observations	9617	9617	9617	9617	9617	9617	9617	9617

Table 9. Public Concern and CSR Activities

This table shows how firms respond to public concern through CSR activities. The sample contains firm-month observations from 2006 and 2020. The dependent variable $I(CSR : Issue)$ takes a value of one if a firm reports a CSR activity addressing an issue ($Issue \in \{Environment,\ Inclusion,\ Education,\ Poverty\}$) in the month, and zero otherwise. $Concern_{t-k,t,Issue}$ is the monthly index of public concern related to an issue ($Issue$), measured over the past k months. Firm fixed effects are included in all regressions. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	I(CSR: Environment)					I(CSR: Inclusion)			
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
$Concern\_t-1,t,Environment$	0.002***				$Concern\_t-1,t,Inclusion$	0.003***			
	[5.04]					[5.87]			
$Concern\_t-3,t,Environment$		0.003***			$Concern\_t-3,t,Inclusion$		0.004***		
		[4.93]					[6.20]		
$Concern\_t-6,t,Environment$			0.003***		$Concern\_t-6,t,Inclusion$			0.005***	
			[4.97]					[6.32]	
$Concern\_t-12,t,Environment$				0.004***	$Concern\_t-12,t,Inclusion$				0.006***
				[4.71]					[5.94]
Firm FE	✓	✓	✓	✓	Firm FE	✓	✓	✓	✓
Adjusted R^2	0.1	0.1	0.1	0.1	Adjusted R^2	0.09	0.09	0.09	0.09
Observations	199296	199296	199296	199296	Observations	199296	199296	199296	199296

	I(CSR: Education)					I(CSR: Poverty)			
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
$Concern\_t-1,t,Education$	0.002***				$Concern\_t-1,t,Poverty$	0.003***			
	[3.26]					[4.20]			
$Concern\_t-3,t,Education$		0.003***			$Concern\_t-3,t,Poverty$		0.003***		
		[3.33]					[3.94]		
$Concern\_t-6,t,Education$			0.003***		$Concern\_t-6,t,Poverty$			0.004***	
			[3.55]					[3.80]	
$Concern\_t-12,t,Education$				0.004***	$Concern\_t-12,t,Poverty$				0.005***
				[3.44]					[4.62]
Firm FE	✓	✓	✓	✓	Firm FE	✓	✓	✓	✓
Adjusted R^2	0.177	0.178	0.178	0.179	Adjusted R^2	0.121	0.121	0.121	0.121
Observations	199296	199296	199296	198427	Observations	199296	199296	199296	198427

Table 10. Strategic Choice of the Method of Addressing Social Issues

This table shows how firms choose the method of addressing environmental and social issues. The sample includes news releases on CSR activities that address four issues: the environment, inclusion, education, and poverty issues within the U.S. *Giving* is an indicator variable that assumes a value of one if a CSR activity is corporate giving, and zero otherwise. $I(Env|Inc)$ is an indicator variable that takes a value of one if a CSR activity pertains to either environmental concerns or issues related to diversity. $CSR\_Concern(M)$, where M is either 3 or 6, indicates the level of public concern for an issue that a CSR activity targets, measured over the preceding M months. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	Giving					
	(1)	(2)	(3)	(4)	(5)	(6)
$I(Env\|Inc)$	-0.463***	-0.456***	0.06	-0.02	0.156	0.06
	[-16.42]	[-15.46]	[0.54]	[-0.16]	[1.14]	[0.41]
$I(Env\|Inc) \times CSR\_Concern(3)$			-0.447***	-0.373***		
			[-5.18]	[-3.99]		
$CSR\_Concern(3)$			0.167***	0.128***		
			[3.80]	[3.05]		
$I(Env\|Inc) \times CSR\_Concern(6)$					-0.529***	-0.440***
					[-4.84]	[-3.76]
$CSR\_Concern(6)$					0.183***	0.136***
					[3.69]	[2.88]
Industry FE	✓		✓		✓	
Firm FE		✓		✓		✓
Adjusted R^2	0.293	0.373	0.298	0.376	0.299	0.377
Observations	10094	9981	10094	9981	10094	9981

Appendix A. Sample description

In this section, I detail the sample. I begin by selecting a set of qualified firms without survivorship bias. Notably, I use the fact that the 1,000 largest U.S. companies account for over 90% of the U.S. stock-market capitalization. I then choose firms whose end-of-fiscal-year market capitalization (represented as 'prcc' × 'csho' in Compustat files) ranks within the top 1,000 among all U.S. firms for any year between 2006 and 2020. I subsequently exclude utility firms (with SIC codes between 4900 and 4999) and financial firms (with SIC codes between 6000 and 6999). This filtering yields 1502 unique firms. Out of these 1502 qualified firms, I collect press releases for 1430 companies throughout the sample period. Table A1 displays the coverage of the final sample.

Table A1. The Coverage of Firms

The table displays both the number of qualified firms and the number of firms for which press releases are collected. "Qualified firms" refer to a subset of U.S. companies that rank among the top 1,000 firms each year and are neither financial nor utility firms. "Covered firms" denote those companies for which we have gathered press releases. In Panel A, Row (A) indicates the number of qualified firms. Row (B) presents the number of covered firms. In Panel B, Row (A) provides the number of firm-year observations for qualified firms throughout the sample period. Row (B) lists the number of firm-year observations for covered firms. Row (C) in either Panel A or Panel B represents the coverage ratio.

Panel A: The coverage of qualified firms								
Year	2006	2007	2008	2009	2010	2011	2012	2013
(A) Qualified firms	696	716	706	719	718	719	722	730
(B) Covered firms	655	680	672	684	688	689	695	702
(C) Coverage: (B)%(A)	0.941	0.950	0.952	0.951	0.958	0.958	0.963	0.962
Year	2014	2015	2016	2017	2018	2019	2020	Total
(A) Qualified firms	711	693	679	683	703	697	730	1502
(B) Covered firms	687	665	655	660	686	682	717	1430
(C) Coverage: (B)%(A)	0.966	0.960	0.965	0.966	0.976	0.978	0.982	0.952
Panel B: The coverage of all firm-year observations of qualified firms								
Year	2006	2007	2008	2009	2010	2011	2012	2013
(A) Qualified firms	1180	1171	1162	1165	1185	1193	1235	1244
(B) Covered firms	1112	1107	1100	1106	1129	1141	1188	1199
(C) Coverage: (B)%(A)	0.942	0.945	0.947	0.949	0.953	0.956	0.962	0.964
Year	2014	2015	2016	2017	2018	2019	2020	Total
(A) Qualified firms	1235	1201	1166	1132	1086	1050	1013	17418
(B) Covered firms	1195	1164	1131	1100	1058	1025	989	16744
(C) Coverage: (B)%(A)	0.968	0.969	0.970	0.972	0.974	0.976	0.976	0.961

Appendix B. Multi-label classification model

In this section, I describe how I categorize press releases into predetermined topics. After perusing several press releases, I identify and classify nine distinct topics. Specifically, the topics include: (1) earnings/performance, (2) mergers and acquisitions (M&A), (3) corporate social responsibility (CSR), (4) changes in leadership, (5) financing activities that encompass equity issuance, debt issuance, or payment of pre-existing debt, (6) directorship, (7) stock repurchase, (8) dividends, and (9) other news pertinent to the business. To develop a deep learning model for this classification task, I utilize transfer learning in natural language processing (NLP). Transfer learning in NLP allows me to build upon prior knowledge accumulated from a vast corpus to address the specific problem I face. There exist various forms of transfer learning, each depending on the source of knowledge utilized. For this project, I employ the Bidirectional Encoder Representations from Transformers (BERT) model from 2018, introduced by Devlin et al. (2019). BERT is a machine learning technique trained on an extensive corpus sourced from the BooksCorpus (800M words) and English Wikipedia (2,500M words). It's widely recognized for producing accurate word representations. Thus, crafting a transfer learning model using a pre-trained BERT model enables me to leverage knowledge from this expansive corpus. To tailor a deep learning model with transfer learning for my specific task, I undertake the following steps.

STEP 1. Building examples

The deep learning model I create is a supervised machine learning model, so I require a set of already labeled examples. I initially select 10,000 random examples and then invest more effort to identify additional CSR-related news. This ensures my model encounters as many CSR articles as possible. The final count of examples reaches 17,824. After removing duplicates, I retain 16,183 unique examples. I then divide these examples into training and test sets, allocating 90% for training and 10% for testing. Additionally, to adjust hyper-parameters, I designate 10% of the training examples as a validation set.

STEP 2. Preprocessing

First, I preprocess titles by removing irrelevant symbols and punctuation marks. Next, I use SpaCy's named entity recognition, trained on the OntoNotes 5 corpus, to replace any identifiable named entities with predetermined tags. For instance, I replace 'Bill Gates' with 'PERSON'. I then tokenize the titles using the BERT uncased tokenizer. Finally, BERT requires specific tokens that characterize documents, namely '[CLS]' and '[SEP]', which I add accordingly.

STEP 3. Generating inputs for the model

I input the tokens prepared in STEP 2 into a pre-trained BERT model to obtain word embeddings. Specifically, I limit the number of tokens to thirty-two and feed them into a version of the BERT model that produces a 768-dimensional embedding for each token. Therefore, an input to my deep learning model is a tensor with a shape of 32 x 768.

STEP 4. Training the model.

I input the embeddings into 9 different Long Short-Term Memory (LSTM) layers with a timestamp of 32. Then, I feed the outputs of the LSTM layers into a combination of dense layers with 'selu' activation functions. All information aggregates in the final dense layer (the output layer) where I obtain nine outputs from sigmoid activation functions. Specifically, I get nine numbers ranging between 0 and 1, which indicate the probability of a title belonging to a particular class. The details of the model are illustrated in Figure B. To compute performance metrics, if a score is greater than 0.5, I classify the title into the respective category. By the model's design, one title can have multiple labels. Of the various metrics evaluating the performance of a classification model, my model achieves 98.3% binary accuracy, 93.7% precision, and 91.3% recall on the test set when applying a threshold of 0.5.

STEP 5. Get predictions on all press releases.

From the model, I receive nine probability-based predictions for each title input. The choice of threshold for classification is at the user's discretion. To maximize the identification of as many CSR news releases as possible, I aim to enhance the model's 'recall' metric. As a result, I set a threshold of 0.2; any input above this threshold receives a label. The binary accuracy, recall, and precision of the model evaluated on the test set appear in Table B2. It's important to note that when I apply a threshold of 0.2, the recall of the CSR class exceeds 97% on the test set. This means I overlook approximately three percent of all CSR news releases when employing the multi-label model.

STEP 6. Get rid of non-CSR related news.

I manually review the articles that the multi-label classification model identifies as CSR news and retain only those that genuinely pertain to CSR. The final results are displayed in Table B3.

Table B1. The count of examples by class

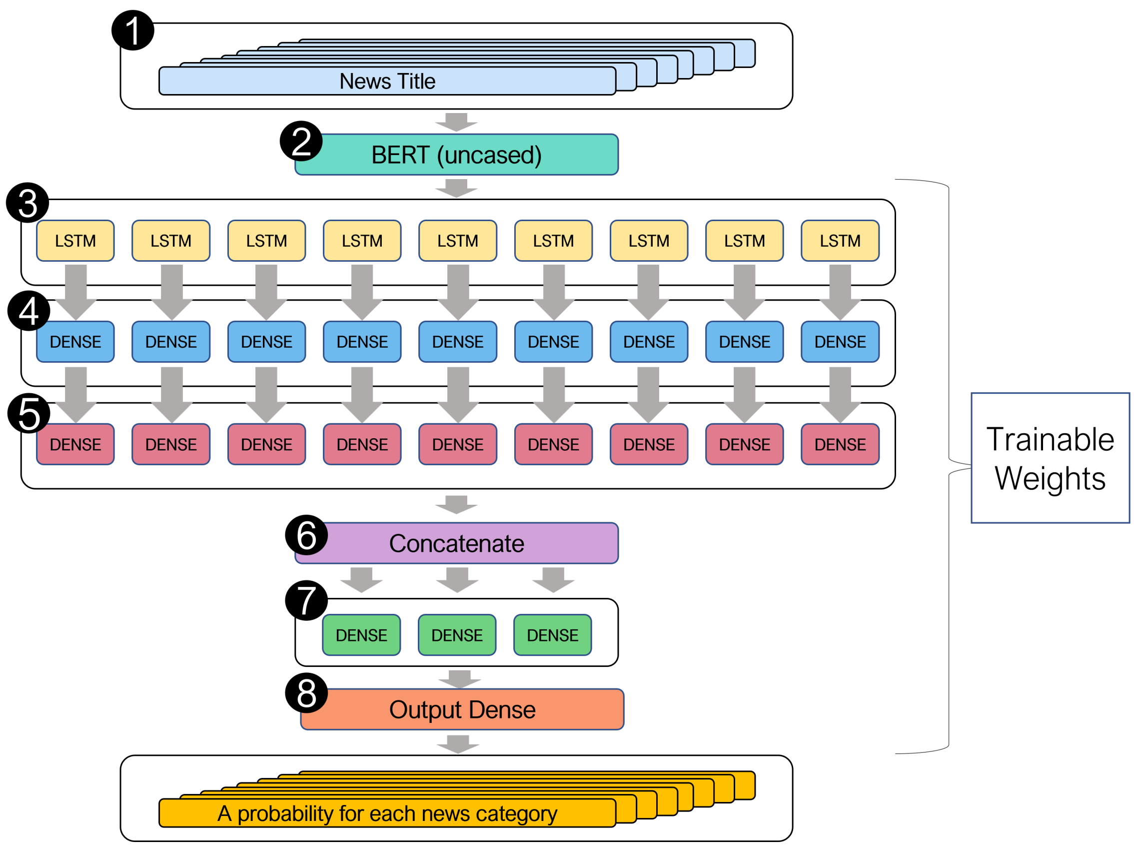
Class	Count
Repurchase	658
Financing	767
Directorship	857
Leadership	964
Dividend	965
M&A	1107
Earnings/Performance	2256
CSR	3522
Other Business	5641
Total	16737

Table B2. Model performance on the test set

	Threshold = 0.5				Threshold = 0.2			
	Accuracy	Recall	Precision	f1 score	Accuracy	Recall	Precision	f1 score
CSR	0.964	0.924	0.908	0.915	0.963	0.971	0.868	0.917
Business	0.939	0.897	0.916	0.906	0.922	0.940	0.842	0.888
Performance	0.985	0.929	0.966	0.947	0.983	0.946	0.938	0.942
M&A	0.987	0.882	0.938	0.909	0.985	0.908	0.885	0.896
Dividend	0.998	0.990	0.981	0.986	0.996	0.990	0.954	0.972
Financing	0.988	0.790	0.970	0.871	0.985	0.840	0.850	0.845
Repurchase	0.998	0.971	0.971	0.971	0.996	1.000	0.919	0.958
Leadership	0.993	0.942	0.942	0.942	0.988	0.971	0.855	0.909
Directorship	0.994	0.903	0.988	0.944	0.993	0.946	0.926	0.936

Figure B. The architecture of the multi-label classification model.

This figure illustrates the deep learning model architecture designed to classify press releases into predefined categories. ❶ The title of a news release is preprocessed before being fed into the model. ❷ A layer of the BERT uncased model converts each token of a sentence into a 768-element vector. The weights of the BERT layer remain untrained. ❸ Each of the nine LSTM layers processes a series of vectors, each representing a token of a sentence, and produces a 32-element vector representing the entire sentence. ❹ Each of the nine linear dense layers, with softmax activation functions, takes a 32-element vector and yields another 32-element vector. All these layers have a 50% dropout rate. ❺ Each of the nine linear dense layers processes a 32-element vector and generates a 9-element vector. ❻ A concatenation layer combines these to produce an 81-element vector. ❼ Each of the three linear dense layers, equipped with selu activation functions, processes an 81-element vector to produce a 9-element vector. These layers have a 0.5 dropout rate. ❽ Three 9-element vectors are concatenated and input to the final output layer, which produces nine probabilities corresponding to the likelihood of a news release belonging to one of the nine categories. The model utilizes the Adam optimizer with an L1 regularizer of 0.0000001. Early stopping with a patience of 4 is implemented, and the loss function is binary cross-entropy.



Internet Appendix

Corporate Social Responsibility Programs and Shareholder Value

Table OA.1. Summary Statistics of Time-Series Public Concern Measures

This table shows the summary statistics of public concern measures. In Panel A, the monthly source-normalized number of articles associated with a social issue (e.g., education) are averaged out across news outlets to produce monthly time series observations $x_{t,p}$, $t \in \{Jan1996, Feb1996, \ldots, Dec2020\}$ for each social issue $p \in \{Environment, Inclusion, Education, Poverty\}$. These values are averaged over the past k months to produce measures $Concern_{t-k,t,p} = \frac{1}{k}\sum_{t-k+1}^{t} x_{\tau,p}$. In Panel B, I introduce another measure where I impose monotone transformation. Specifically, a monthly measure of public concern in Panel B is the log of one plus the average of $x_{t,p}$ over a specific number of months k measured in a month t, that is $CSR\_Concern(k,p) = log(1 + \frac{1}{k}\sum_{t-k+1}^{t} x_{\tau,p})$.

Panel A: Public Concern Measures

	N	Mean	Std. Dev.	Min	P25	Median	P75	Max
$Concern_{t-1,t,Environment}$	180	2.47	0.64	1.08	2.03	2.43	2.88	4.54
$Concern_{t-1,t,Inclusion}$	180	2.18	0.62	0.84	1.81	2.10	2.45	4.72
$Concern_{t-1,t,Education}$	180	1.88	0.55	0.67	1.47	1.86	2.23	3.86
$Concern_{t-1,t,Poverty}$	180	2.01	0.55	0.64	1.58	1.99	2.38	3.59
$Concern_{t-3,t,Environment}$	180	2.46	0.55	1.22	2.10	2.40	2.84	3.71
$Concern_{t-3,t,Inclusion}$	180	2.17	0.50	1.09	1.86	2.12	2.49	4.11
$Concern_{t-3,t,Education}$	180	1.87	0.46	0.83	1.49	1.89	2.20	2.84
$Concern_{t-3,t,Poverty}$	180	2.01	0.48	1.02	1.63	1.97	2.35	3.14
$Concern_{t-6,t,Environment}$	180	2.45	0.50	1.35	2.17	2.45	2.91	3.38
$Concern_{t-6,t,Inclusion}$	180	2.14	0.43	1.15	1.91	2.11	2.43	3.85
$Concern_{t-6,t,Education}$	180	1.87	0.43	0.83	1.49	1.87	2.19	2.79
$Concern_{t-6,t,Poverty}$	180	2.00	0.45	1.25	1.64	1.99	2.28	3.03
$Concern_{t-12,t,Environment}$	180	2.42	0.47	1.36	2.15	2.47	2.79	3.27
$Concern_{t-12,t,Inclusion}$	180	2.11	0.39	1.08	1.94	2.15	2.31	3.08
$Concern_{t-12,t,Education}$	179	1.86	0.40	1.17	1.52	1.84	2.21	2.66
$Concern_{t-12,t,Poverty}$	179	1.98	0.42	1.25	1.63	1.95	2.19	2.84

Panel B: Log-Transformed Public Concern Measures

		p =Environment		p =Inclusion		p =Education		p =Poverty	
	N	Mean	Std	Mean	Std	Mean	Std	Mean	Std
$CSR\_Concern(1,p)$	180	1.2281	0.1925	1.1381	0.1834	1.0403	0.1618	1.0865	0.1570
$CSR\_Concern(3,p)$	180	1.2293	0.1633	1.1407	0.1582	1.0426	0.1503	1.0879	0.1399
$CSR\_Concern(6,p)$	180	1.2273	0.1524	1.1360	0.1469	1.0413	0.1460	1.0867	0.1321

Table OA.2. Firm Characteristics and CSR activities

This table shows the relationship between firm characteristics and CSR activities. The sample contains firm-year observations in a period from 2006 to 2020. The dependent variable $I(CSR)_{t+1}$ is an indicator that takes one if a firm reports a CSR activity in the next year, and zero otherwise. Similarly, the dependent variable $I(Issue)_{t+1}$, $Issue \in \{Environment, Inclusion, Education, Poverty\}$, is an indicator that takes one if a firm reports a CSR activity addressing an issue ($Issue$) in the next year, and zero otherwise. The dependent variable $I(Edu|Pov)_{t+1}$ is an indicator that takes one if a firm reports an activity tackling education or poverty issues in the next year, zero otherwise. Avg_roa_2y is the two-year average return on assets (ROA). $chair\_CEO$ is an indicator variable that takes one when a CEO is the chairman of the board, and zero otherwise. $log(\#director)$ is the logarithm of one plus the number of directors on a company's board. $frac\_ind$ is the proportion of independent directors on the board. $size$ is the logarithm of total assets. $cash$ is cash and cash equivalent divided by total assets. ml is market leverage. mb is market-to-book ratio. All regressions include firm- and year-fixed effects. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	(1)	(2)	(3)	(4)	(5)	(6)
	$I(CSR)_{t+1}$	$I(Environment)_{t+1}$	$I(Inclusion)_{t+1}$	$I(Education)_{t+1}$	$I(Poverty)_{t+1}$	$I(Edu\|Pov)_{t+1}$
Avg_roa_2y	-0.043	-0.008	-0.022	0.033	0.023	0.036
	[-0.783]	[-0.237]	[-0.944]	[1.137]	[0.660]	[0.891]
$chair\_CEO$	0.009	-0.01	0.004	0.015**	0.01	0.018*
	[0.763]	[-1.029]	[0.566]	[2.007]	[1.099]	[1.848]
$log(\#director)$	0.002	-0.007	-0.005	0.004	-0.005	-0.008
	[0.163]	[-0.700]	[-0.639]	[0.595]	[-0.649]	[-0.823]
$frac\_ind$	0.022	-0.014	-0.008	0	0.012	0.009
	[0.839]	[-0.689]	[-0.498]	[0.033]	[0.674]	[0.459]
$size$	0.046***	0.011	0.010*	0.013**	0.016**	0.025***
	[3.962]	[1.378]	[1.759]	[2.052]	[2.208]	[2.909]
$cash$	-0.028	0.037	0.019	-0.006	-0.026	0.001
	[-0.595]	[1.381]	[1.030]	[-0.237]	[-0.961]	[0.026]
ml	-0.003	0.029	-0.026	-0.008	0.02	0.006
	[-0.078]	[1.049]	[-1.289]	[-0.384]	[0.728]	[0.187]
mb	0.004	0.001	-0.004**	-0.002	0.002	-0.001
	[0.950]	[0.301]	[-2.348]	[-0.860]	[0.643]	[-0.193]
Firm FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Adjusted R^2	0.49	0.385	0.298	0.452	0.418	0.464
Observations	12873	12873	12873	12873	12873	12873

Table OA.3. Timing of EPS Announcement and CSR News Releases

This table shows whether the timing of EPS announcements and CSR news releases correlate depending on the content of EPS reports. In the first four regressions, the dependent variable is a dummy variable that assumes a value of one if a firm discloses a CSR activity within the 30 days preceding the earnings announcement. For the subsequent four regressions, the dependent variable is an indicator, assigned a value of one when a firm announces a CSR activity within the 30 days following the earnings announcement and zero otherwise. $EPSsurprise$ is EPS surprises. EPS surprises are measured by actual quarterly EPS minus the median forecast where the median forecast is found from forecasts made between 2 and 15 days prior to the earnings announcement if available. Otherwise, the median forecast is found from forecasts made between 2 and 30 days before the earnings announcement. The measure is scaled by the share price of the firm, which is an observation available on the day closest to the EPS announcement day between five days before the announcement and three days before the event, inclusively. $I(EPSsurprise < 0)$ is an indicator variable that takes one if $EPSsurprise$ is negative, zero otherwise. Size, market-to-book ratio, market leverage, and cash holdings based on information from a year before are included in all regressions. Firm and year-fixed effects are included in all regressions. t-statistics based on standard errors clustered at the firm level are shown below the coefficient estimates. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level.

	CSR news within 30 days before EPS news				CSR news within 30 days after EPS news			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$EPSsurprise$	0.0062	0.0131	0.009	-0.0036	-0.0731	-0.0537	-0.0972	-0.0995
	[0.0709]	[0.1515]	[0.0935]	[-0.0377]	[-0.7656]	[-0.5619]	[-0.9633]	[-0.9898]
$I(EPSsurprise < 0)$			0.0004	-0.0009			-0.0008	-0.0023
			[0.1348]	[-0.3112]			[-0.2534]	[-0.7571]
size	0.0168***	0.0200***	0.0159***	0.0198***	0.0168***	0.0212***	0.0160***	0.0214***
	[5.4296]	[4.6457]	[5.0329]	[4.5518]	[5.3575]	[4.2795]	[5.0216]	[4.3298]
cash	-0.0570***	-0.0349**	-0.0613***	-0.0369**	-0.0415**	-0.0163	-0.0479**	-0.0201
	[-3.2742]	[-2.0558]	[-3.4544]	[-2.1351]	[-2.2912]	[-0.9352]	[-2.5443]	[-1.1066]
ml	-0.0165	-0.0059	-0.0177	-0.006	-0.0118	0.0063	-0.0124	0.0078
	[-1.0650]	[-0.3618]	[-1.1249]	[-0.3655]	[-0.7248]	[0.3732]	[-0.7743]	[0.4644]
mb	-0.0002	0.0032**	-0.0003	0.0031**	0.0015	0.0048***	0.0011	0.0044***
	[-0.1462]	[2.4249]	[-0.2672]	[2.2689]	[1.1072]	[3.2284]	[0.8316]	[3.0271]
Firm FE	✓	✓	✓	✓	✓	✓	✓	✓
Year FE		✓		✓		✓		✓
Adjusted R^2	0.234	0.245	0.228	0.239	0.231	0.242	0.226	0.237
Observations	43925	43925	41374	41374	43925	43925	41374	41374

Peer Effects on CSR Activities

Peer effects are known to exist in corporate decisions such as financial policies (Leary and Roberts (2014), Grennan (2019)). However, it is an empirical question whether such peer effects exist in the implementation of CSR initiatives. If investors know that firms engage in CSR because of their peers, they will reflect the information into stock prices. In this section, I examine the existence of peer effects in CSR initiatives.

To examine peer effects in CSR policies, I draw on the literature that uses instrumental variable (IV) approaches to navigate the reflection problem discussed by Manski (1993). Specifically, I adopt the approach in Leary and Roberts (2014) and Grennan (2019). The two papers exploit the fact that peer firms' idiosyncratic stock returns could be viewed a potential source of exogenous variation in peer firms' financial decisions. Leary and Roberts (2014) relate peer firms' idiosyncratic volatility to peer firm leverage while Grennan (2019) associates it with peer firm dividend policy. The idea that idiosyncratic risk can have implications on dividend policy predicates on the findings by Hoberg and Prabhala (2009), whose work finds that an increase in risk is a significant driver behind a reduction in dividend payments of firms in their sample period.

I rely on a similar idea. Findings in Masulis and Reza (2014) indicate that managers might use cash for corporate giving when they can otherwise return it to shareholders in the form of dividends. This implies that firms can direct financial resources to implement CSR instead of distributing them to their shareholders. Combining the findings of Hoberg and Prabhala (2009) and Masulis and Reza (2014), it leads to a conclusion that if firm risk can affect dividend policies, then firm risk should also have implications on CSR activities. Moreover, the link between firm risk and CSR activities should be similar to the link between firm risk and dividend payments documented in the literature. Therefore, the hypothesis for the first stage regression is that peer firms' idiosyncratic volatility is negatively related to the probability of peer firms conducting CSR activities.

To test the hypothesis, I construct peer firms' idiosyncratic volatility following Grennan (2019). Peer groups are defined by the firms having the same the three-digit SIC code. I first calculate peer firms' idiosyncratic return. To that end, each firm's daily raw return is disentangled into market-level, industry-level and idiosyncratic firm-level shocks following Campbell et al. (2001). Daily excess returns are computed using Treasury bill rates. Then, I subtract from the daily excess return the industry-wide component of returns constructed by value-weighted daily excess returns within a peer group with a weight being a market capitalization at the end of the previous year. This process results in firms' idiosyncratic returns. The idiosyncratic risk of a firm i in a peer group p in a year t is then calculated with more than 100 daily (d) observations for each year as follows.

$$idio\_vol_{i,p,t} = \sum_{d \in t} idio\_ret^2_{i,p,t,d} \tag{11}$$

Finally, peer idiosyncratic risk ($peer\_IVol$) is constructed by taking the average idiosyncratic risk of the peer firms excluding own-firm idiosyncratic risk. The mean value of peer idiosyncratic risk is 0.1278 and standard deviation is 0.1857.

In the first stage regression, I use peer idiosyncratic risk measured at the beginning of the year as a main instrumental variable. However, following the literature I also include contemporaneous peer firm idiosyncratic volatility as another IV.

The dependent variable is the peer participation in CSR, or peer influence in the domain of CSR. I define peer influence by the fraction of peer firms that take on CSR activities. $PeerCSR$ represents the fraction of peer firms that take on CSR activities in a given year, encompassing all initiatives beyond the scope of the four specified issues. $PeerIssue$, ($Issue \in \{Environment,\ Inclusion,\ Poverty,\ Education\}$) is the fraction of peers that address an issue ($Issue$) in a given year.

The results of the first stage regressions are presented in the table below (**Table OA.4**). The results show that peer idiosyncratic volatility is a strong predictor of the CSR participation of the peers. Column (a) indicates that one standard deviation increase in peer idiosyncratic volatility measured at the beginning of the year ($Peer\_IVol_{t-1}$) leads to a decrease in the fraction of peers implementing CSR in that year by 0.0294. Among the four types of CSR activities, environmental CSR and philanthropic CSR are more heavily affected by idiosyncratic volatility.

In the second stage regression, the fitted value from the first regression is used as a peer influence free of endogeneity issues. With the peer influence, I examine how it affects own-firm's decision making on CSR activities. I run the following regression with peer influence measured at the beginning of the year (t), allowing the variable is observable by the time the firm I makes a decision.

$$I(CSR)_{i,p,t+1} = peer\_influence_{i,p,t} + X'_{i,p,t}\gamma + H'_{-i,p,t}\xi + \mu_p + \phi_{t+1} + \epsilon_{i,p,t+1} \tag{12}$$

In the above regression specification, subscript i is a firm, t is a year, p is a peer group. $I(CSR)_{i,p,t+1}$ is an indicator variable that takes a value one when a firm i reports a CSR activity in year $t+1$. $peer\_influence_{i,p,t}$ is the fitted value from a first stage regression. $X_{i,p,t}$ is a set of firm characteristics that are size, market leverage, market-to-book ratio, cash, and the three board characteristics included in the previous analyses. $H_{-i,p,t}$ is peer firm averages of the same firm characteristics. μ_p, ϕ_{t+1}, and $\epsilon_{i,t+1}$ are fixed effects and the error term.

The results in **Table OA.4** show that not all CSR activities exhibit peer effects. Two CSR types, environmental and philanthropic CSR, show statistically significant presence of peer effects. Also, the sign of the coefficients tell us that the peer effects in general work in a way that it induces a firm to take on CSR when its peers take on CSR projects. In column (c), I find that one standard deviation increase in the fraction of peers that report CSR activities leads to an increase in the probability of own-firm's conducting CSR in the next year by 0.047. [25] In column (g), one standard deviation increase in the fraction of peers that implement philanthropic CSR in year t increases the probability of own-firm's releasing CSR news related to philanthropic activities in year $t+1$ by 0.083. The economic magnitude of peer effects on the environmental and philanthropic CSR is not negligible.

[25] The mean value of peer influence for all CSR is 0.252894 with a standard deviation of 0.239116. The mean value of peer influence for CSR related to the environment is 0.0537463 with a standard deviation of 0.1258942. The mean value of peer influence for CSR on inclusive society is 0.0470042 with a standard deviation of 0.1139507 . The mean value of peer influence for CSR related to education is 0.0628419 with a standard deviation of 0.1215725 . The mean value of peer influence for philanthropic CSR is 0.0911462 with a standard deviation of 0.1646439.

Table OA.4. CSR Peer Effects

This table shows the peer effects on CSR activities. The sample contains firm-year observations of sample firms between 2006 and 2020. Peers are defined as firms that share the same first three-digit SIC code. Each column presents the results from a first-stage regression followed by the corresponding second-stage regression results. $PeerCSR$ represents the fraction of peer firms that take on CSR activities in a given year, encompassing all initiatives beyond the scope of the four specified issues. $PeerIssue$, ($Issue \in \{Environment,\ Inclusion,\ Poverty,\ Education\}$) is the fraction of peers that address an issue ($Issue$) in a given year. The instrumental variable for peer influence in the first-stage regressions is peer firms' idiosyncratic volatility ($Peer\_IVol$), constructed following Grennan (2019). In the second stage regression, an indicator variable for a firm's CSR activities is regressed on peer influence which is a fitted value from the first-stage regression in the same column and measured a year before, allowing the fitted value to be observable by the time a firm makes a decision. Firm-level controls include $size$, $cash$, ml, mb, $log(\#director)$, $chair\_ceo$, and $frac\_ind$ measured a year before the dependent variable. Peer-average controls are peers' average $size$, $cash$, ml, mb, $log(\#director)$, $chair\_ceo$, and $frac\_ind$ measured a year before the dependent variable. 'Year FE' and 'Industry FE' stand for year-fixed effects and industry-fixed effects, respectively. Standard errors are reported in square brackets and clustered at the firm level. */**/*** indicate statistical significance at the 10%, 5%, and 1% level.

1st Stage	$PeerCSR_t$		$PeerEnvironment_t$		$PeerInclusion_t$		$PeerPoverty_t$		$PeerEducation_t$	
$Peer\_IVol_{t-1}$	-0.158***	-0.133***	-0.082***	-0.074***	-0.051***	-0.029***	-0.100***	-0.082***	-0.034***	-0.023***
	[0.014]	[0.015]	[0.008]	[0.009]	[0.007]	[0.008]	[0.011]	[0.011]	[0.008]	[0.008]
$Peer\_IVol_t$		-0.063***		-0.018***		-0.055***		-0.046***		-0.028***
		[0.012]		[0.007]		[0.006]		[0.009]		[0.007]
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.173	0.175	0.032	0.032	0.073	0.079	0.041	0.043	0.017	0.019
2nd Stage	$I(CSR)_{t+1}$		$I(Environment)_{t+1}$		$I(Inclusion)_{t+1}$		$I(Poverty)_{t+1}$		$I(Education)_{t+1}$	
Peer influence$_t$	0.174	0.095	0.377**	0.399**	0.174	0.001	0.506**	0.375*	-0.072	0.089
	[0.135]	[0.140]	[0.189]	[0.186]	[0.299]	[0.163]	[0.219]	[0.195]	[0.488]	[0.375]
Firm-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Peer average controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.236	0.236	0.16	0.16	0.131	0.131	0.184	0.184	0.158	0.158
Observations	11381	11375	11381	11375	11381	11375	11381	11375	11381	11375

Table OA.5. The List of Representative Articles

Social Issue	Title	News date
Environment	Getting serious about plastic	2017-11-07
Environment	The tragic reason seabirds keep mistaking ocean plastic for food	2016-11-10
Environment	Recycling needs to be everyone's priority	2016-09-20
Environment	Air Pollution Can Be Deadly for Seniors	2017-12-29
Environment	OZONE, POLLUTION LEVELS HEADING HIGHER TODAY	2005-08-02
Environment	Recycling isn't enough – the world's plastic pollution crisis is only getting worse	2020-09-29
Environment	Plastic Pollution That Creates Global Concern	2020-02-10
Environment	Sustainability groups discuss need for more involvement to reach zero waste	2019-04-23
Environment	The Eco-Conscious Pay to Ease Guilt	2006-12-10
Environment	The Fracker's Guide to a Greener World	2012-11-12
Environment	Global Warming and Mt. Kilimanjaro	2009-12-07
Environment	Carbon Caps Are the Best Policy	2009-03-24
Environment	Pollution From Ozone Is a Lot More Harmful To Us Than It Looks	1998-06-22
Environment	Hot Air	1997-06-27
Environment	Devasted by drought	2012-09-25
Environment	Imperfect 'cap-and-trade' is best option to fight warming	2009-11-17
Environment	Prince Charles wages a 'green' campaign online	2009-05-06
Environment	Meat and the Planet	2006-12-27
Environment	Why I'm Giving $1 Billion for the Planet	2018-11-01
Environment	An Ounce of Science Versus a Ton of Cure	2018-03-13
Environment	Shed a Tear for the Reefs	2017-03-19
Environment	World economy, carbon free by 2050	2017-03-24
Environment	Deadly Combination; Humans and Climate Change Go Way Back'	2016-06-21
Environment	Teaching the Truth About Climate Change	2015-10-11
Environment	The G.O.P. Can't Ignore Climate Change	2014-05-07
Environment	Life After Land	2011-07-19
Environment	Colorless Green Ideas	2007-02-23
Environment	Blinding Ourselves in Space	2007-01-21
Environment	Climate Change Gets Real For Americans	2012-12-26
Environment	A year of extreme weather, and no reprieve in sight	2012-12-26
Environment	Americans waste 150,000 tons of food each day – equal to a pound per person	2018-04-18
Environment	Your Recycling Gets Recycled, Right? Maybe, or Maybe Not	2018-05-29
Environment	More than 8.3 billion tons of plastics made: Most has now been discarded	2017-07-19
Environment	World's Oceans Clogged by Millions of Tons of Plastic Trash	2015-02-12
Environment	New era of 'super fires' as climate change triggers hotter, drier weather	2016-05-11
Environment	Here's What We Know about Wildfires and Climate Change	2017-10-13
Environment	Extreme heat and wildfires made worse by climate change, say scientists	2018-07-28
Environment	Have We Passed the Acid Test?	2018-05-02
Environment	Global warming: Improve economic models of climate change	2014-04-04
Environment	We're almost out of time: The alarming IPCC climate report and what to do next	2018-10-16
Environment	Climate change made Australia's devastating fire season 30% more likely	2020-03-04
Environment	Climate change made European heatwave up to 3°C hotter	2019-08-02
Environment	Droughts, heatwaves and floods: How to tell when climate change is to blame	2018-07-30
Environment	Extreme weather explicitly blamed on humans for the first time	2017-12-19
Environment	Global warming: Shareholders must vote for climate-change mitigation	2016-02-10
Environment	Legal threat exposes gaps in climate-change planning	2017-08-31
Environment	Pinning extreme weather on climate change is now routine and reliable science	2018-07-30
Environment	Climatologists to physicists: your planet needs you	2015-04-07
Environment	Waste Crisis: Americans Create 3x More Waste Than Global Average	2019-07-03
Environment	Study: Air pollution causes 200,000 early deaths each year in the U.S.	2013-08-29

Social Issue	Title	News date
Inclusion	Education gap threatens students' economic future	2002-12-18
Inclusion	Racial education gap debated ; Speakers call on schools to make greater effort	2002-07-09
Inclusion	Professors discuss LGBTQ issues with students	2015-10-20
Inclusion	Public School Reform Would Close Racial Gap in Education, Authors Say	2003-11-20
Inclusion	Cross Country: Tech Workers and Asians Against Racial Preferences'	2019-10-26
Inclusion	Obama Needs to Take a Stand on Race and Other Issues	2008-08-28
Inclusion	We must disarm racism and hate	2020-06-17
Inclusion	Integration Now and Forever	2018-03-30
Inclusion	Racism Without Racists	2008-10-05
Inclusion	Black Lives Matter Is Democracy in Action	2017-10-22
Inclusion	Google employee spreadsheet alleges wide pay gap for women	2017-09-13
Inclusion	Why women earn less	2008-06-06
Inclusion	The business case for diversity in the workplace is now overwhelming	2019-04-29
Inclusion	Diversity And Inclusion Matters To The Workforce Of The Future	2018-05-09
Inclusion	Why LGBT Employees Need Workplace Allies	2013-06-20
Inclusion	Moving from commitment to action on LGBTI equality	2019-01-23
Inclusion	How can I help my company increase workplace diversity? Ask HR	2019-02-18
Inclusion	The Black-white wealth gap left Black households more vulnerable	2020-12-08
Inclusion	Yes, social justice and discrimination were driving issues for Latino voters in 2020	2020-11-06
Inclusion	Unequal Opportunity: Race and Education	1998-03-01
Inclusion	Investors are the biggest losers when women and minority entrepreneurs don't get startup money	2019-10-07
Inclusion	Why Don't More Women Start Businesses?	2017-06-11
Inclusion	There Are Few Minority Entrepreneurs, And They Rarely Get Funding	2013-10-16
Inclusion	Part-Time Penalty Hits Working Mothers	2014-08-21
Inclusion	Gender Imbalance in the Lab	2014-05-24
Inclusion	Motherhood Still a Cause Of Pay Inequality	2012-06-13
Inclusion	A Gender Bias In Film Reviewing	2018-07-18
Inclusion	Job Interviews Without Gender	2018-01-07
Inclusion	What life is like as a transgender woman	2020-06-22
Inclusion	Gay marriage ruling reflects new dimensions of freedom	2015-06-29
Inclusion	Coming of Age and Coming Out	2019-05-26
Inclusion	Marching in Washington; Gay People Demonstrate, In Pride and in Fear	1993-05-02
Inclusion	Why Minorities Have So Much Trouble Accessing Small Business Loans	2018-01-22
Inclusion	LGBTQ community isn't waiting for Equality Act to pass	2021-11-15
Inclusion	How Will the American Workforce Change?	2015-12-31
Inclusion	Study: Race, poverty define education gap ; Schools plan to reduce disparity in achievement	2005-08-16
Inclusion	Spending said to lag in poor, minority schools	2005-12-22
Inclusion	Gallup: Workplace Bias Still Prevalent	2006-02-01
Inclusion	U.S. high school dropout rate reaches record low, driven by improvements among Hispanics, blacks	2014-10-02
Inclusion	The Surprising Ways The Gender Wage Gap Affects Families	2015-11-05
Inclusion	K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities	2018-03-22
Inclusion	STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity	2021-04-01
Inclusion	Women and Men in STEM Often at Odds Over Workplace Equity	2018-01-09

Social Issue	Title	News date
Poverty	Study: Race, poverty define education gap	2005-08-16
Poverty	Spending said to lag in poor, minority schools	2005-12-22
Poverty	Food Stamps Shouldn't Pay for Junk	2018-04-10
Poverty	The Missing Element to Beat Poverty	2019-05-30
Poverty	Winning The War On Poverty	2019-04-05
Poverty	America's Deep Poverty Problem	2018-01-25
Poverty	Growing Up Poor in America	2016-10-30
Poverty	In the War on Poverty, a Dogged Adversary	2013-12-18
Poverty	Are the Poor Suffering From Hunger Anymore?	2003-02-23
Poverty	Researchers 'surprised' by what happened when low-income moms received regular cash payments	2022-01-25
Poverty	Hunger in America could get worse as supply chains tighten	2022-01-21
Poverty	Hunger lingers for millions of underemployed, low-income Americans	2021-12-14
Poverty	Safety net for poor unravels; Poverty is increasing, but problem; often overlooked in political debate	2004-10-14
Poverty	Food insecurity among certain households big	2021-09-09
Poverty	Child Poverty in South Dakota: A Statistical Profile	2006-12-01
Poverty	The War Isn't Over; Despite Washington claims, poverty still gripping Phila.	2018-07-30
Poverty	Reducing hunger and poverty - school breakfast pays off.	2018-03-11
Poverty	Students Shouldn't Have to Choose Between Books and Food	2016-02-28
Poverty	Efforts to feed thousands of low-income children barely make a dent rising child hunger	2015-07-24
Poverty	Poverty, not uneven funding, explains the achievement gapr	2018-12-07
Poverty	Hunger doesn't take a vacation	2015-05-27
Poverty	Poverty tied to school performance	2019-09-12
Poverty	Majority Believe There Will be More Poor Americans Four Years from Now	2005-01-11
Poverty	Born Into Poverty and Obesity	2016-03-23
Poverty	Ashley Zhang: When good health is not always a choice	2017-03-06
Poverty	Research spotlights the grim effect of poverty on education	2015-05-13
Poverty	Youth from low-income family risk their health for success	2015-07-14
Poverty	Homeless youth on the rise, with state funding in question	2016-02-14
Poverty	Homeless students arise from many different situations	2016-12-27
Poverty	Grow economy by shrinking poverty	2018-11-09
Poverty	Behind the numbers: Millions seeking a path out of poverty	2018-09-12
Poverty	More children living in poverty now than during recession	2015-07-21
Poverty	Older, Suburban and Struggling, 'Near Poor' Startle the Census	2011-11-18
Poverty	Report: Rural Poverty In America Is 'An Emergency'	2018-05-31
Poverty	Poverty and Opportunity: Begin with Facts	2014-01-28
Poverty	The U.N. Looks At Extreme Poverty In The U.S., From Alabama To California	2017-12-12
Poverty	Over 48 million Americans live in poverty	2014-10-16
Poverty	Growth Has Been Good for Decades. So Why Hasn't Poverty Declined?	2014-06-04
Poverty	Poverty in America: Why Can't We End It?	2012-07-28
Poverty	Federal report: U.S. hunger remains at highest levels in 15 years	2010-11-16

Social Issue	Title	News date
Education	Higher-ed investment essential to region	2018-12-09
Education	How underfunding schools really hurts kids	2012-07-14
Education	The education gap	2002-08-12
Education	University of Chicago Targets Its Inequality	2014-10-02
Education	Analysts: Evidence-based school funding model working, needs more investment	2019-03-28
Education	Pro: Investing in education is key to having top-notch system	2019-01-08
Education	Are colleges ready for STEM students?	2011-11-18
Education	Why Science Majors Change Their Mind	2011-11-06
Education	Hacking the STEM syllabus	2018-12-20
Education	Envisioning STEM education for all	2018-12-19
Education	3 reasons Florida schools should focus on STEM education	2015-12-29
Education	Top 10 education policy wishes	2012-12-20
Education	A Rising Call to Promote STEM Education and Cut Liberal Arts Funding	2016-02-22
Education	Engineering education	2017-09-01
Education	MfA President ponders STEM education crisis, solutions	2017-04-20
Education	Let's confront teacher-quality question in education reform	2017-12-27
Education	Why so many teachers need a second job to make ends meet	2016-12-18
Education	Editorial: STEM teachers may need a premium to stay in class	2016-11-16
Education	Shortages have schools creating future math teachers: Apprentice program trains students	2006-12-28
Education	Teachers can also benefit from school choice	2003-11-11
Education	Study: Race, poverty define education gap ; Schools plan to reduce disparity in achievement	2005-08-16
Education	Spending said to lag in poor, minority schools	2005-12-22
Education	The Diminishing Returns of a College Degree	2017-06-05
Education	College Aid Hiding in Plain Sight	2020-07-01
Education	THE NATION; The View From America's Stranded Public Schools	1988-12-18
Education	Education Does Reduce Inequality	2015-04-10
Education	The Diminishing Returns of a College Degree	2017-06-05
Education	A High-Tech Rebirth From Higher Ed's Ruins	2017-01-23
Education	College Aid Hiding in Plain Sight	2020-07-01
Education	The Hidden Inequality in Schools	2020-01-30
Education	Higher Education and the Opportunity Gap	2013-10-08
Education	School environments can be toxic. Why and how they must change.	2022-01-10
Education	National high school graduation rates at historic high, but disparities still exist	2014-04-28
Education	The True Cost of High School Dropouts	2012-01-25
Education	With Innovation, Colleges Fill the Skills Gap	2017-06-07
Education	5 key findings on what Americans and scientists think about science	2015-01-29
Education	Higher Education Today: Innovative Approaches for College Financing	2013-10-04
Education	Teacher Quality Widely Diffused, Ratings Indicate	2012-02-24
Education	Training of Teachers Is Flawed, Study Says	2011-07-21
Education	Teach Your Teachers Well	2016-01-13
Education	Skills in the digital age - How should education systems evolve?	2016-10-05
Education	The Rising Cost of Not Going to College	2014-02-11
Education	How Teachers Are Using Technology at Home and in Their Classrooms	2013-02-28
Education	Three Reasons College Matters for Social Mobility	2015-02-06
Education	Not just college: Technical education as a pathway to the middle class	2016-04-01
Education	How Higher Education Can Improve Economic Mobility in the United States	2014-10-30
Education	U.S. students' academic achievement still lags that of their peers in many other countries	2017-02-15

References

Abuzov, Rustam, Will Gornall, and Ilya A Strebulaev, 2022, The value of privacy and the choice of limited partners by venture capitalists, *Available at SSRN 4235337* .

Acemoglu, Daron, Ufuk Akcigit, and William Kerr, 2016, Innovation network, Tech. rep., National Bureau of Economic Research.

Ahmadpoor, Mohammad and Benjamin F Jones, 2017, The dual frontier: Patented inventions and prior scientific advance, *Science* 357, 583–587.

Andonov, Aleksandar, Rob MMJ Bauer, and KJ Martijn Cremers, 2017, Pension fund asset allocation and liability discount rates, *The Review of Financial Studies* 30, 2555–2595.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *The journal of finance* 61, 259–299.

Angrist, Joshua D and Guido W Imbens, 1995, Two-stage least squares estimation of average causal effects in models with variable treatment intensity, *Journal of the American statistical Association* 90, 431–442.

Anselmsson, Johan, Niklas Vestman Bondesson, and Ulf Johansson, 2014, Brand image and customers' willingness to pay a price premium for food brands, *Journal of Product & Brand Management* 23, 90–102.

Ardia, David, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht, 2020, Climate change concerns and the performance of green versus brown stocks, *SSRN Electronic Journal* .

Arts, Sam, Jianan Hou, and Juan Carlos Gomez, 2021, Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures, *Research Policy* 50, 104144.

Badawi, Adam B and Frank Partnoy, 2022, Social good and litigation risk, *Harv. Bus. L. Rev.* 12, 315.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, 2016, Measuring economic policy uncertainty, *The Quarterly Journal of Economics* 131, 1593–1636.

Barber, Brad M and Ayako Yasuda, 2017, Interim fund performance and fundraising in private equity, *Journal of Financial Economics* 124, 172–194.

Barillas, Francisco and Jay Shanken, 2016, Which alpha?, *The Review of Financial Studies* 30, 1316–1338.

Barrot, Jean-Noël, 2017, Investor horizon and the life cycle of innovative firms: Evidence from venture capital, *Management Science* 63, 3021–3043.

Bedford, Anna, Le Ma, Nelson Ma, and Kristina Vojvoda, 2021, Patenting activity or innovative originality?, *Accounting & Finance* 61, 4191–4207.

Begenau, Juliane, Pauline Liang, and Emil Siriwardane, 2023, The rise in alternatives .

Bellstam, Gustaf, Sanjai Bhagat, and J. Anthony Cookson, 2021, A text-based analysis of corporate innovation, *Management Science* 67, 4004–4031.

Bénabou, Ronald and Jean Tirole, 2010, Individual and corporate social responsibility, *Economica* 77, 1–19.

Bernstein, Shai, Xavier Giroud, and Richard R Townsend, 2016, The impact of venture capital monitoring, *The Journal of Finance* 71, 1591–1622.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *The Journal of business* 45, 444–455.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky, 2022, When is TSLS actually LATE?, Tech. rep.

Bonelli, Maxime, 2023, The adoption of artificial intelligence by venture capitalists, *SSRN Electronic Journal* .

Breitzman, Anthony and Patrick Thomas, 2015, The emerging clusters model: A tool for identifying emerging technologies across multiple patent systems, *Research policy* 44, 195–205.

Brody, Shaked, Uri Alon, and Eran Yahav, 2021, How attentive are graph attention networks?, *arXiv preprint arXiv:2105.14491* .

Brown, Jeffrey R and David W Wilcox, 2009, Discounting state and local pension liabilities, *American Economic Review* 99, 538–542.

Buchner, Axel and Niklas F Wagner, 2017, Rewarding risk-taking or skill? the case of private equity fund managers, *Journal of Banking & Finance* 80, 14–32.

Cahan, Steven F., Chen Chen, Li Chen, and Nhut H. Nguyen, 2015, Corporate social responsibility and media coverage, *Journal of Banking & Finance* 59, 409–422.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *The Journal of Finance* 56, 1–43.

Capelle-Blancard, Gunther and Aurélien Petit, 2017, Every little helps? ESG news and stock market reaction, *Journal of Business Ethics* 157, 543–565.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *The Journal of finance* 52, 57–82.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *The Journal of finance* 56, 2431–2456.

Chaplinsky, Susan and Swasti Gupta-Mukherjee, 2016, Investment risk allocation and the venture capital exit market: Evidence from early stage investing, *Journal of banking & finance* 73, 38–54.

Cheng, Ing-Haw, Harrison Hong, and Kelly Shue, 2013, Do managers do good with other people's money?, Tech. rep., National Bureau of Economic Research.

Choi, Youngjae, Sanghyun Park, and Sungjoo Lee, 2021, Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data, *Scientometrics* 126, 5431–5476.

Cochrane, John H, 2005, The risk and return of venture capital, *Journal of financial economics* 75, 3–52.

Cohen, Lauren, Karl Diether, and Christopher Malloy, 2013, Misvaluing innovation, *The Review of Financial Studies* 26, 635–666.

Cooper, Michael J, Huseyin Gulen, and Michael J Schill, 2008, Asset growth and the cross-section of stock returns, *the Journal of Finance* 63, 1609–1651.

Cotei, Carmen and Joseph Farhat, 2018, The m&a exit outcomes of new, young firms, *Small Business Economics* 50, 545–567.

Coval, Joshua D and Tobias J Moskowitz, 2001, The geography of investment: Informed trading and asset prices, *Journal of political Economy* 109, 811–841.

Cumming, Douglas and Na Dai, 2010, Local bias in venture capital investments, *Journal of empirical finance* 17, 362–380.

Cunningham, Colleen, Florian Ederer, and Song Ma, 2021, Killer acquisitions, *Journal of Political*

Economy 129, 649–702.

Daniel, Kent and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *The Journal of Finance* 61, 1605–1643.

Dass, Nishant, Vikram Nanda, and Steven Chong Xiao, 2017, Truncation bias corrections in patent data: Implications for recent research on innovation, *Journal of Corporate Finance* 44, 353–374.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2018, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* .

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2019, BERT: Pre-training of deep bidirectional transformers for language understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186.

Dhaliwal, Dan S., Oliver Zhen Li, Albert Tsang, and Yong George Yang, 2011, Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting, *The Accounting Review* 86, 59–100.

Di Giuli, Alberta and Leonard Kostovetsky, 2014, Are red or blue companies more likely to go green? politics and corporate social responsibility, *Journal of Financial Economics* 111, 158–180.

Dimson, Elroy, Oğuzhan Karakaş, and Xi Li, 2015, Active ownership, *Review of Financial Studies* 28, 3225–3268.

Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, 2007, Valuation in over-the-counter markets, *The Review of Financial Studies* 20, 1865–1900.

Eberhart, Allan C, William F Maxwell, and Akhtar R Siddique, 2004, An examination of long-term abnormal stock returns and operating performance following R&D increases, *The journal of finance* 59, 623–650.

Edmans, Alex, 2011, Does the stock market fully value intangibles? employee satisfaction and equity prices, *Journal of Financial economics* 101, 621–640.

Eisfeldt, Andrea L and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance* 68, 1365–1406.

El Ghoul, Sadok, Omrane Guedhami, Chuck C.Y. Kwok, and Dev R. Mishra, 2011, Does corporate social responsibility affect the cost of capital?, *Journal of Banking & Finance* 35, 2388–2406.

Ellis, Jesse, Leonardo Madureira, and Shane Underwood, 2020, The causal effects of proximity on investment: Evidence from flight introductions, *Journal of Financial and Quantitative Analysis* 55, 1978–2004.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel, 2020, Hedging climate change news, *The Review of Financial Studies* 33, 1184–1216.

Érdi, Péter, Kinga Makovi, Zoltán Somogyvári, Katherine Strandburg, Jan Tobochnik, Péter Volf, and László Zalányi, 2013, Prediction of emerging technologies based on analysis of the us patent citation network, *Scientometrics* 95, 225–242.

Fama, Eugene F and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of financial economics* 33, 3–56.

Fama, Eugene F and Kenneth R French, 1996, Multifactor explanations of asset pricing anomalies, *The journal of finance* 51, 55–84.

Fama, Eugene F and Kenneth R French, 2006, Profitability, investment and average returns, *Journal* *of financial economics* 82, 491–518.

Fama, Eugene F and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of financial economics* 116, 1–22.

Fama, Eugene F and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist, 2020, What is a patent worth? evidence from the us patent "lottery", *The Journal of Finance* 75, 639–682.

Ferrell, Allen, Hao Liang, and Luc Renneboog, 2016, Socially responsible firms, *Journal of Financial Economics* 122, 585–606.

Fitza, Markus, Sharon F Matusik, and Elaine Mosakowski, 2009, Do vcs matter? the importance of owners on performance variance in start-up firms, *Strategic Management Journal* 30, 387–404.

Fitzgerald, Tristan, Benjamin Balsmeier, Lee Fleming, and Gustavo Manso, 2021, Innovation search strategy and predictable returns, *Management science* 67, 1109–1137.

Flammer, Caroline, 2013, Corporate social responsibility and shareholder reaction: The environmental awareness of investors, *The Academy of Management Journal* 56, 758–781.

Flammer, Caroline, 2015, Does corporate social responsibility lead to superior financial performance? a regression discontinuity approach, *Management Science* 61, 2549–2568.

Flammer, Caroline, 2021, Corporate green bonds, *Journal of Financial Economics* 142, 499–516.

Fleming, Lee and Olav Sorenson, 2004, Science as a map in technological search, *Strategic management journal* 25, 909–928.

Freeman, R Edward, 2010, Strategic management: A stakeholder approach, Cambridge university press.

Friedman, Milton, 1970, The social responsibility of business is to increase its profits, *New York Times Magazine* 32, 122–126.

Fu, Fangjian, 2009, Idiosyncratic risk and the cross-section of expected stock returns, *Journal of financial Economics* 91, 24–37.

Furukawa, Takao, Kaoru Mori, Kazuma Arino, Kazuhiro Hayashi, and Nobuyuki Shirakawa, 2015, Identifying the evolutionary process of emerging technologies: A chronological network analysis of world wide web conference sessions, *Technological Forecasting and Social Change* 91, 280–294.

Gaule, Patrick, 2018, Patents and the success of venture-capital backed startups: Using examiner assignment to estimate causal effects, *The Journal of Industrial Economics* 66, 350–376.

Gifford, Sharon, 1997, Limited attention and the role of the venture capitalist, *Journal of business venturing* 12, 459–482.

Giot, Pierre, Ulrich Hege, and Armin Schwienbacher, 2014, Are novice private equity funds risk-takers? evidence from a comparison with established funds, *Journal of Corporate Finance* 27, 55–71.

Goetzmann, William N, Alok Kumar, et al., 2005, Why do individual investors hold under-diversified portfolios?, Tech. rep.

Gompers, Paul and Josh Lerner, 1999, What drives venture capital fundraising?

Gompers, Paul A, Will Gornall, Steven N Kaplan, and Ilya A Strebulaev, 2020, How do venture capitalists make decisions?, *Journal of Financial Economics* 135, 169–190.

González-Uribe, Juanita, 2020, Exchanges of innovation resources inside venture capital portfolios, *Journal of Financial Economics* 135, 144–168.

Goss, Allen and Gordon S. Roberts, 2011, The impact of corporate social responsibility on the cost of bank loans, *Journal of Banking & Finance* 35, 1794–1810.

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

Griffin, John M and Michael L Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *The Journal of Finance* 57, 2317–2336.

Gu, Feng, 2005, Innovation, future earnings, and market efficiency, *Journal of Accounting, Auditing & Finance* 20, 385–418.

Gu, Lifeng, 2016, Product market competition, r&d investment, and stock returns, *Journal of Financial Economics* 119, 441–455.

Guest, Paul M, 2009, The impact of board size on firm performance: evidence from the uk, *The European Journal of Finance* 15, 385–404.

Ha-Brookshire, Jung E and Pamela S Norum, 2011, Willingness to pay for socially responsible products: case of cotton apparel, *Journal of consumer Marketing* 28, 344–353.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The nber patent citation data file: Lessons, insights and methodological tools.

Hamilton, James T., 2003, All the News That's Fit to Sell, Princeton University Press.

Hamilton, Will, Zhitao Ying, and Jure Leskovec, 2017, Inductive representation learning on large graphs, *Advances in neural information processing systems* 30.

Hart, Oliver and Luigi Zingales, 2017, Companies should maximize shareholder welfare not market value, *SSRN Electronic Journal* .

Hartzmark, Samual M. and Abigail B. Sussman, 2019, Do investors value sustainability? a natural experiment examining ranking and fund flows, *The Journal of Finance* 74, 2789–2837.

Hellmann, Thomas and Manju Puri, 2002, Venture capital and the professionalization of start-up firms: Empirical evidence, *The journal of finance* 57, 169–197.

Higgins, Matthew J and Daniel Rodriguez, 2006, The outsourcing of r&d through acquisitions in the pharmaceutical industry, *Journal of financial economics* 80, 351–383.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2013, Innovative efficiency and stock returns, *Journal of financial economics* 107, 632–654.

Hirshleifer, David, Po-Hsuan Hsu, and Dongmei Li, 2018, Innovative originality, profitability, and stock returns, *The Review of Financial Studies* 31, 2553–2605.

Hoberg, Gerard and Gordon Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *The Review of Financial Studies* 23, 3773–3811.

Hoberg, Gerard and Nagpurnanand R. Prabhala, 2009, Disappearing dividends, catering, and risk, *Review of Financial Studies* 22, 79–116.

Hochberg, Yael V, Alexander Ljungqvist, and Yang Lu, 2010, Networking as a barrier to entry and the competitive supply of venture capital, *The Journal of Finance* 65, 829–859.

Hochberg, Yael V and Joshua D Rauh, 2013, Local overweighting and underperformance: Evidence from limited partner private equity investments, *The Review of Financial Studies* 26, 403–451.

Hochberg, Yael V, Carlos J Serrano, and Rosemarie H Ziedonis, 2018, Patent collateral, investor commitment, and the market for venture lending, *Journal of Financial Economics* 130, 74–94.

Hochreiter, Sepp and Jürgen Schmidhuber, 1997, Long short-term memory, *Neural computation* 9, 1735–1780.

Höck, André, Christian Klein, Alexander Landau, and Bernhard Zwergel, 2020, The effect of environmental sustainability on credit risk, *Journal of Asset Management* 21, 85–93.

Hong, Harrison and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hou, Kewei and Roger K Loh, 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* 121, 167–194.

Jaffe, Adam B, Manuel Trajtenberg, and Michael S Fogarty, 2000, The meaning of patent citations: Report on the nber/case-western reserve survey of patentees.

Jia, Ning and Dan Wang, 2017, Skin in the game: General partner capital commitment, investment behavior and venture capital fund performance, *Journal of Corporate Finance* 47, 110–130.

Jones, Charles M and Owen A Lamont, 2002, Short-sale constraints and stock returns, *Journal of Financial Economics* 66, 207–239.

Kamepalli, Sai Krishna, Raghuram Rajan, and Luigi Zingales, 2020, Kill zone, Tech. rep.

Kaplan, Sarah and Keyvan Vakili, 2014, The double-edged sword of recombination in breakthrough innovation, *Strategic Management Journal* 36, 1435–1457.

Kaplan, Steven N and Antoinette Schoar, 2005, Private equity performance: Returns, persistence, and capital flows, *The journal of finance* 60, 1791–1823.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, 2021, Measuring technological innovation over the long run, *American Economic Review: Insights* 3, 303–320.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf, 2014, Entrepreneurship as experimentation, *Journal of Economic Perspectives* 28, 25–48.

Kim, Gabjo and Jinwoo Bae, 2017, A novel approach to forecast promising technology through patent analysis, *Technological Forecasting and Social Change* 117, 228–237.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *The Quarterly Journal of Economics* 132, 665–712.

Korteweg, Arthur and Morten Sorensen, 2017, Skill and luck in private equity performance, *Journal of Financial Economics* 124, 535–562.

Krüger, Philipp, 2015, Corporate goodness and shareholder wealth, *Journal of Financial Economics* 115, 304–329.

Kyebambe, Moses Ntanda, Ge Cheng, Yunqing Huang, Chunhui He, and Zhenyu Zhang, 2017, Forecasting emerging technologies: A supervised learning approach through patent analysis, *Technological Forecasting and Social Change* 125, 236–244.

Leary, Mark T. and Michael R. Roberts, 2014, Do peer firms affect corporate financial policy?, *The Journal of Finance* 69, 139–178.

Lee, Changyong, Ohjin Kwon, Myeongjung Kim, and Daeil Kwon, 2018, Early identification of emerging technologies: A machine learning approach using multiple patent indicators, *Technological Forecasting and Social Change* 127, 291–303.

Lee, Charles MC, Stephen Teng Sun, Rongfei Wang, and Ran Zhang, 2019, Technological links and predictable returns, *Journal of Financial Economics* 132, 76–96.

Lerner, Josh and Amit Seru, 2022, The use and misuse of patent data: Issues for finance and beyond, *The Review of Financial Studies* 35, 2667–2704.

Leung, Woon Sau, Kevin P Evans, and Khelifa Mazouz, 2020, The r&d anomaly: Risk or mispricing?, *Journal of Banking & Finance* 115, 105815.

Levy, Haim, 1978, Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio, *The American Economic Review* 68, 643–658.

Li, Dongmei, 2011, Financial constraints, r&d investment, and stock returns, *The Review of Financial Studies* 24, 2974–3007.

Lintner, John, 1975, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Stochastic optimization models in finance, Elsevier, 131–155.

Lu, Lina, Matt Pritsker, Andrei Zlate, Kenechukwu Anadu, and James Bohn, 2019, Reach for yield by us public pension funds, *FRB Boston Risk and Policy Analysis Unit Paper No. RPA* , 19–2.

Ma, Song, 2021, Technological obsolescence, Tech. rep.

Makri, Marianna, Michael A Hitt, and Peter J Lane, 2010, Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions, *Strategic management journal* 31, 602–628.

Manski, Charles F., 1993, Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60, 531.

Margolis, Joshua D., Hillary Anger Elfenbein, and James P. Walsh, 2009, Does it pay to be good...and does it matter? a meta-analysis of the relationship between corporate social and financial performance, *SSRN Electronic Journal* .

Masulis, Ronald W and Rajarishi Nahata, 2011, Venture capital conflicts of interest: Evidence from acquisitions of venture-backed firms, *Journal of Financial and Quantitative Analysis* 46, 395–430.

Masulis, Ronald W. and Syed Walid Reza, 2014, Agency problems of corporate philanthropy, *Review of Financial Studies* 28, 592–636.

Merton, Robert C, 1973, An intertemporal capital asset pricing model, *Econometrica: Journal of the Econometric Society* , 867–887.

Merton, Robert C et al., 1987, A simple model of capital market equilibrium with incomplete information .

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, 2013, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems* 26.

Mittal, Vrinda, 2022, Desperate capital breeds productivity loss: Evidence from public pension investments in private equity, *Available at SSRN* .

Mohan, Nancy and Ting Zhang, 2014, An analysis of risk-taking behavior for public defined benefit pension plans, *Journal of Banking & Finance* 40, 403–419.

Nahata, Rajarishi, 2008, Venture capital reputation and investment performance, *Journal of financial economics* 90, 127–151.

Nanda, Ramana and Matthew Rhodes-Kropf, 2013, Investment cycles and startup innovation, *Journal of financial economics* 110, 403–418.

Newey, Whitney K and Kenneth D West, 1986, A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix .

Nguyen, Bang Dang and Kasper Meisner Nielsen, 2010, The value of independent directors: Evidence from sudden deaths, *Journal of financial economics* 98, 550–567.

Pástor, Ľuboš, Robert F. Stambaugh, and Lucian A. Taylor, 2021, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.

Pennacchi, George and Mahdi Rastad, 2011, Portfolio allocation for public pension funds, *Journal* *of Pension Economics & Finance* 10, 221–245.

Pew Research Center, 2018, State public pension funds' investment practices and performance: 2016 data update, Tech. rep.

Pierrakis, Yannis and George Saridakis, 2017, Do publicly backed venture capital investments promote innovation? differences between privately and publicly backed funds in the uk venture capital market, *Journal of Business Venturing Insights* 7, 55–64.

Reimers, Nils and Iryna Gurevych, 2019, Sentence-bert: Sentence embeddings using siamese bert-networks.

Riedl, Arno and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *The Journal of Finance* 72, 2505–2550.

Rotolo, Daniele, Diana Hicks, and Ben R Martin, 2015, What is an emerging technology?, *Research policy* 44, 1827–1843.

Serafeim, George and Aaron Yoon, 2021, Stock price reactions to ESG news: The role of ESG ratings and disagreement, *SSRN Electronic Journal* .

Sharpe, William F, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The journal of finance* 19, 425–442.

Sorensen, Morten, 2004, How smart is smart money? an empirical two-sided matching model of venture capital, *Unpublished working paper. University of Chicago* .

Sorenson, Olav and Toby E Stuart, 2001, Syndication networks and the spatial distribution of venture capital investments, *American journal of sociology* 106, 1546–1588.

Tian, Xuan, 2011, The causes and consequences of venture capital stage financing, *Journal of Financial Economics* 101, 132–159.

Tian, Xuan and Tracy Yue Wang, 2014, Tolerance for failure and corporate innovation, *The Review of Financial Studies* 27, 211–255.

Veličković, Petar, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio, 2017, Graph attention networks, *arXiv preprint arXiv:1710.10903* .

Von Neumann, John and Oskar Morgenstern, 2007, Theory of games and economic behavior, Theory of games and economic behavior, Princeton university press.

Watzinger, Martin and Monika Schnitzer, 2019, Standing on the shoulders of science .

Zhang, X Frank, 2006, Information uncertainty and stock returns, *The journal of Finance* 61, 105–137.

Zhou, Yuan, Fang Dong, Yufei Liu, Zhaofu Li, JunFei Du, and Li Zhang, 2020, Forecasting emerging technologies using data augmentation and deep learning, *Scientometrics* 123, 1–29.

Zhou, Yuan, Fang Dong, Yufei Liu, and Liang Ran, 2021, A deep learning framework to early identify emerging technologies in large-scale outlier patents: An empirical study of cnc machine tool, *Scientometrics* 126, 969–994.