Venture Capital and Scientists' Selection into Entrepreneurship

Xuelai Li*

Link to Latest Version

October 2025

Abstract

This paper examines the causal effect of venture capital (VC) on scientists' selection into entrepreneurship. I collect a novel historical dataset of U.S. scientists in the 1960s and track their business formation activities. I leverage the reform of the "prudent man" rule under the Employee Retirement Income Security Act (ERISA) as a natural experiment that positively shocks the supply of VC. I use large language models to further exploit the cross-sectional variation in how scientists' work specialties rely on tangible versus intangible capital. I show that scientists' business formation increased by 91% post ERISA. Effects are stronger for those with intangible specialties and those working in the private sector. These scientists were not marginal entrants but had higher wages and were named inventors on patents. I rationalize the results and quantify the effect of VC on alleviating financial constraints through an occupational choice model. I show that the individual-level effects ultimately facilitate the growth of intangible industries at the county level.

Keywords: Venture Capital, Entrepreneurship, Intangible Capital, Innovation

*Department of Finance, Imperial College London, x.li20@imperial.ac.uk. I am indebted to Ramana Nanda, Cláudia Custódio, Rajkamal Iyer, and Savitar Sundaresan for all their guidance. I also thank Jamie Coen, Patrick Coen, Harsha Dutta, Stephen Hansen, Jonathan Haskel, Ziang Li, Lu Liu, Clara Martínez-Toledano, Adrien Matray, Darcy Pu, Tarun Ramadorai, Ailsa Röell, Magdalena Rola-Janicka, Claudia Schaffranka (discussant), Juhana Siljander, Tong Yu, and seminar participants at the Wharton Innovation Doctoral Symposium, BSE Summer Forum, Scientific Workshop on Productivity, HEC Paris Finance PhD Workshop, and Imperial College London for their valuable comments and suggestions. All errors are my own.

1 Introduction

Technology entrepreneurship is widely recognized as a critical source for economic growth. By fostering innovation, creating employment opportunities, and fueling competitiveness, technology entrepreneurship stimulates broader market development and improves productivity (Aghion and Howitt, 1992; Schumpeter, 1983).

Venture capital (VC) has long been recognized as a driver of technology entrepreneurship (Nanda and Rhodes-Kropf, 2017; Howell, 2017). Macro-level evidence suggests that regions receiving greater VC inflows exhibit stronger economic growth and innovation (Chen and Ewens, 2025; Fehder, Hausman, and Hochberg, 2025; Kortum and Lerner, 2000; Samila and Sorenson, 2011). Skilled labor forming startups is especially important in the context of technology entrepreneurship (Acemoglu, Akcigit, Alp, Bloom, and Kerr, 2018; Akcigit and Kerr, 2018; Christensen, 2011).

Despite the recognized importance of VC in financing innovation, we have limited causal evidence of the impact of VC on the entrepreneurial entry decisions of skilled labor at the individual level. In other words, do promising projects by skilled labor go unfunded without VC? If so, what type of projects are left unfunded?

Answering these questions is empirically challenging for three main reasons. First, the definition of skilled labor is ambiguous, so the data sample of skilled labor can vary across contexts. Second, entrepreneurial entry decisions at the individual level remain largely unexamined due to a lack of systematic data. Typically, only those who start businesses are observed in archival databases, while constructing a comprehensive dataset on *would-be* entrepreneurs is empirically demanding. Third, there are few historical instances of exogenous shocks to the supply of VC. VC endogenously flows to places where there are opportunities and local entrepreneurial ecosystems, and such places are likely to take off even in the absence of VC funding.

To overcome these challenges, I construct a novel panel of U.S. scientists who were active in the 1960s as a representative sample of skilled labor.[1] I compile a cross-sectional

[1]Scientists in this paper are defined as scientific and technical personnel engaged in research and development.

snapshot of their educational backgrounds and work experiences, and then expand it into panel data by linking the scientists to business registration data to observe their selection into entrepreneurship. I choose this time period because the data I use is anonymized after 1972, and only the earlier data contains both names and residence locations, which allows me to match scientists to business registration data.

I leverage the 1979 reform of the "prudent man" rule under the Employee Retirement Income Security Act (ERISA) in the U.S. as an exogenous shock to the supply of VC. This ERISA reform implemented by the Department of Labor relaxed pension fund allocation restrictions and substantially increased the pool of capital available to VC firms (Kortum and Lerner, 2000; Gompers, 1994). Prior to this reform, VC firms had difficulty raising funds because the "prudent man" rule, as one of the fiduciary rules of ERISA, restricted pension fund investments in higher-risk assets such as small firm equity. Existing research leverages this policy change to examine the effect of VC in international settings and at the country level (Gornall and Strebulaev, 2021), while this paper focuses on how financial constraints affect individual decisions regarding entrepreneurial entry across tangible and intangible work specialties.

I exploit the cross-sectional variation in scientists' work specialties by classifying the specialties according to their reliance on tangible assets. VC seeks scalability and outsized returns, which are usually more common in businesses that rely more on intangible capital.[2] Moreover, as a specialized financial intermediary, VC shares risks while closely monitoring entrepreneurs to mitigate moral hazard problems associated with intangible investments (Beck, Döttling, Lambert, and van Dijk, 2023). Therefore, scientists with intangible work specialties are more likely to be affected by the expansion of VC. Importantly, these scientists did not select their specialties in anticipation of future VC inflows, as the U.S. VC market was negligible in the 1960s[3].

The main finding contains four sets of results. First, following the ERISA reform,

[2]There are several measures of intangibles at the firm level, using R&D expenses and selling, general, and administrative (SG&A) expenses (Crouzet and Eberly, 2021; Eisfeldt and Papanikolaou, 2013, 2014; Peters and Taylor, 2017). In this study, instead of at firm level, I measure the reliance of intangible capital at the scientific specialty level.

[3]Fewer than one hundred early-stage investment deals per year, according to Venture Economics.

scientists became more likely to start businesses. Business formation rate rose from 0.78% in the seven years before the reform to 1.49% in the seven years after, representing a 91% increase. Moreover, the result remains robust when (i) relying on continuous measures of tangibility, (ii) excluding computer science–related scientists to isolate the effect from the concurrent rise of personal computing and associated intangible business opportunities, and (iii) dropping Delaware to account for the legal structure of business registration.

Second, although all scientists responded to the ERISA reform, those with intangible specialties had a business formation rate that was relatively 0.05 percentage points higher than that of scientists with tangible specialties. Given the pre-ERISA business formation rate of 0.78%, this effect represents a 6.41% increase relative to the baseline.

Third, the effects are primarily driven by scientists working in the private sector. Private scientists with intangible specialties had a business formation rate that was relatively 0.18 percentage points higher than that of private scientists with tangible specialties. Given the pre-ERISA business formation rate of 1.22% by private scientists, this effect represents a 14.8% increase relative to the baseline. By contrast, university scientists hardly responded to the shock. This is consistent with an entrepreneurial spawning mechanism.[4] The evidence underscores the role of VC in incentivizing inventors to commercialize innovations outside of corporate boundaries.

Fourth, the effects are concentrated among private scientists with higher wages and prior patenting activity. Responsiveness to VC supply increases with wage quartiles: the higher the wage group, the stronger the response. Among private scientists with at least one patent before ERISA, those with intangible specialties had a business formation rate that was 0.74 percentage points higher than scientists with tangible specialties. This indicates a 25.5% relative effect of starting a business with a baseline rate of 2.90% before ERISA for the private scientists who filed at least one patent. By contrast, the difference is only 0.17% among scientists without prior patents. This result suggests that VC disproportionately encourages high-quality technology entrepreneurship.

[4] Entrepreneurial spawning process is that individuals leave large bureaucratic companies and become entrepreneurs (Gompers, Lerner, and Scharfstein, 2005).

To rationalize the empirical findings, I extend the occupation choice model developed by Evans and Jovanovic (1989) to capture the financial constraints and entrepreneurial entry decisions. Scientists choose between wage employment and entrepreneurship. Entrepreneurial production requires investment in physical and intangible capital. Scientists finance businesses through funds raised against wealth and by pledging physical capital. Since only physical capital is pledgeable, financial constraints bind more tightly for scientists with intangible specialties. Venture capital relaxes these constraints by increasing the share of funds that can be raised against wealth, thereby making intangible scientists more likely to enter entrepreneurship. The model shows that scientists with higher productivity are more likely to be financially constrained, which is consistent with the empirical results that productive scientists are more responsive to the VC shock. Moreover, the model quantifies the role of VC in relaxing financial constraints. Using the data from my analysis and the literature, this framework shows that VC relaxes financial constraints by 69.2%.

At the county-industry level, I document that VC facilitates the expansion of intangible industries. Employment and establishment exhibited substantial growth in intangible industries after the ERISA reform. Moreover, the effect is more salient in counties with ex post VC presence. These aggregate trends are consistent with the individual-level evidence on scientists.

Overall, these findings demonstrate that VC significantly enhances the rate of business formation among scientists, particularly those with intangible work specialties. Far from being marginal entrants, these scientists often held patents and were recognized as productive inventors. Moreover, VC contributes to broader economic growth by enabling the expansion of the intangible economy.

Related Literature. This paper contributes to several strands of literature in entrepreneurship and financial intermediation. First, this paper speaks to the VC and technology entrepreneurship literature, which highlights the role of VC-backed firms in driving IPOs (Lerner and Nanda, 2020) and underscores the importance of monitoring, staged financing, and value-added services (Bernstein, Giroud, and Townsend, 2016;

Gompers, 1995; Korteweg and Sensoy, 2023). Gornall and Strebulaev (2021) use G7 members as a comparison group and show that 65% of top public firms would not have been founded in the U.S. without the ERISA reform. However, the mechanisms through which VC incentivizes skilled labor to start a business are not yet fully understood. This paper shows causal evidence of the effect of VC on scientists' business formation decisions and explores the heterogeneity within the effects. Consistent with Babina, Bernstein, and Mezzanotti (2023), who find that the Great Depression contributed to the decline of technological entrepreneurship and accelerated the shift of innovation toward larger firms, I find that VC deregulation had the opposite effect, fostering entrepreneurial spawning as employees left established firms to start businesses. The emergence of spinouts from established technology firms further proves the role of VC in enhancing market competitiveness, a key driver of long-term innovation and economic growth (Cunningham, Ederer, and Ma, 2021; Ma, 2025).

Second, my results provide insight into the literature on financial intermediation and small business financing. Financial intermediaries play a key role in enhancing liquidity and reducing information asymmetries. Prior research highlights that banks are critical sources of financing for small and medium-sized enterprises (SMEs), which often rely on debt (Robb and Robinson, 2014; Kerr and Nanda, 2009; Nanda and Nicholas, 2014) and home equity (Corradin and Popov, 2015; Kerr, Kerr, and Nanda, 2022). Most of the research on financial intermediation centers around banks. Literature has shown that the intangibles of large firms are primarily associated with cash flow-based lending or unsecured debt, whereas tangibles are associated with asset-based lending or secured debt (Benmelech, Kumar, and Rajan, 2022, 2024). Moreover, banks have difficulty in assessing the quality of an innovation as part of the intangible capital. As a result, a funding gap emerges—one that VC is better positioned to fill. Departing from Hellmann, Lindsey, and Puri (2008), who study how banks use VC investments to build lending relationships, I show that VC, as a special type of financial intermediary, is particularly good at financing technology startups and complements banks.

Finally, the findings contribute to the broader discussion on the role of financial intermediaries in economic growth. The share of intangible assets in firms' capital stock and the aggregate economy has increased markedly in recent decades (Corrado, Hulten, and Sichel, 2009; Crouzet and Eberly, 2021; Eisfeldt and Papanikolaou, 2014; Haskel and Westlake, 2017). Prior research shows that bank financing can support intangible investment when patents serve as collateral, facilitating access to debt capital (Mann, 2018; Morse, 2024). Relatedly, Hochberg, Serrano, and Ziedonis (2018) document the use of patent-backed assets in venture lending. This paper offers a complementary perspective: VC is particularly well suited for financing intangible assets due to their non-rival nature and scalability (Crouzet, Eberly, Eisfeldt, and Papanikolaou, 2022). The expansion of the VC market has played a critical role in supporting the growth of intangible industries.

The rest of this paper is organized as follows. Section 2 provides an overview of the historical context of VC in small business financing. Section 3 describes the data sources and presents descriptive statistics on the scientists included in the analysis. Section 4 examines the reduced-form relationship between the VC supply and business formation and explores the heterogeneity. Section 5 rationalizes the empirical results and quantifies the effect of VC in relaxing financial constraints through an occupational choice model. Section 6 investigates the effects of VC at macro level. Section 7 concludes.

2 Historical Context

The financing landscape for technology entrepreneurship remained largely informal until the advent of VC in 1959, marked by the establishment of Draper, Gaither & Anderson (DGA), the first VC firm structured as a limited partnership. DGA's investment strategy laid the groundwork for private capital investment, emphasizing four key criteria: "(1) companies offering unique products or services, (2) substantially developed offerings with predictable commercialization timelines and costs, (3) a clearly identifiable market, and (4) the presence of or access to qualified management." Similarly, Greylock's 1965 offering memorandum underscored a preference for speculative startups characterized by innovative products, processes, or technologies (Nicholas, 2019).

However, raising capital for new ventures posed significant challenges because of the limited investment avenues available for entrepreneurs. Traditional sources of funding, such as Small Business Investment Companies (SBICs)[5], were off-limits to those unwilling to accept government loans. Additionally, institutional investors, such as pension funds, were constrained by regulatory frameworks like the "prudent man" rule, which prohibited investments in higher-risk assets, including VC (Zock, 1980). This left individual investors as a potential source of funding, but this route presented its own challenges. The volatility of personal wealth, stemming from events such as divorce or death, created issues regarding the valuation of invested capital and could result in protracted disputes over the worth of early-stage ventures. Consequently, the difficulty of securing funding in this era was compounded by a complex interplay of regulatory constraints and the inherent risks of dealing with individual investors. By the mid-1970s, there were no more than about 30 fairly substantial VC firms nationwide. Even the more established VCs, such as Greylock and Venrock, managed relatively small investment pools by modern standards (Nicholas, 2019).

The absence of institutional investors as limited partners and regulatory constraints on pension fund investments further restricted the growth of the VC industry, leaving early-stage startups with limited funding opportunities. Before the ERISA reform in 1979, the "prudent man" rule deterred many pension managers from allocating capital into VC funds, as investing in small business securities can be seen as imprudent. ERISA set the fiduciary requirement imposed on private pension funds, according to which a manager must discharge their duty "with the care, skill, prudence, and diligence under the circumstances then prevailing that a prudent man acting in a like capacity and familiar with such matters would use in the conduct of an enterprise of a like character and with like aims." A fiduciary must protect investors by continually monitoring. The fiduciary requirements imply that investing in small business securities can be of high

[5]SBICs are private funds licensed by the U.S. Small Business Administration (SBA) that invest in small firms using their own capital and SBA-guaranteed debt, which allows them to borrow at favorable terms and expand financing capacity.

risk. Moreover, ERISA was overseen by both the Treasury and the Department of Labor at that time, which imposed unnecessarily complex administrative requirements.

In August 1978, President Jimmy Carter proposed the ERISA reorganization plan to Congress, and it was approved in October. [6] The Treasury was to have statutory authority for the minimum standards, while the Department of Labor (DOL) was to have statutory authority for the fiduciary obligations.

In June 1979, the DOL explicitly clarified the fiduciary requirement in a federal register (Figure A1), allowing fund managers to invest capital in venture funds as part of their total portfolio. This reform significantly increased the supply of capital to VC funds, as shown in Figure A2. The fundraising patterns were mirrored in the investments by VC into small firms (Kortum and Lerner, 2000). In a similar spirit, the staggered adoption of "prudent man" rules, prompted by the 1994 Uniform Prudent Investor Act, also increases capital commitments to the local VC industry (González-Uribe, 2020).

The composition of limited partners in VC funds changed significantly due to the ERISA reform. Prior to the ERISA reform, the limited partners of VC funds were evenly distributed among industrial corporations, insurance companies, foundations, and individuals. But by 1984, pension funds had become the single most important source of VC funds (Florida and Kenney, 1988). It is important to note that ERISA regulations do not apply to state pension funds, as these funds are governed by state laws rather than federal regulations. State pension funds typically adhere to more conservative investment strategies, prioritizing fixed income and public equities.[7] While ERISA exclusively affects private pension funds, these funds generally exhibit greater allocations to VC compared to state pension funds.

[6] Message to Congress Transmitting Reorganization Plan No. 4 of 1978.

[7] However, some of the largest state pension funds (e.g., CalPERS, CalSTRS, NYSCRF, Texas TRS) were pioneers in investing in alternative assets since the 1980s.

3 Historical Data

3.1 Scientists and Engineers

At the heart of this paper is the comprehensive database of historical scientific and technical personnel of the U.S. that I assembled. To comprehensively understand the state of U.S. scientific and technical personnel in the 1960s, I collected individual-level data from two sources: the National Register of Scientific and Technical Personnel (NRSTP) from the National Archives and the American Men of Science (AMS). I then tracked their business formation activities from 1970 onward.

There are two main reasons for collecting data on scientists in the 1960s instead of earlier or later years. First, the ERISA shock was unanticipated during this period, as the U.S. VC market was still in its early stages. Therefore, scientists did not select their work specialties based on future financing opportunities. Even if they had such intentions, they would have likely chosen tangible specialties that were more easily financed by banks. Second, the NRSTP data ends in 1972; subsequent versions are anonymized, preventing the linkage between scientists and business registrations.

3.1.1 National Register of Scientific and Technical Personnel

I retrieved the NRSTP dataset from the National Archives Access to Archival Databases. The NRSTP was initially created by the National Science Foundation (NSF) to identify specialized professionals for national emergencies, but once the data's utility for statistical analysis was recognized, its primary function shifted toward providing a key source of statistical information on scientific and engineering personnel.[8] It provided critical data for developing science policy and supplied information to Congress and government agencies.

The NRSTP records professionals in various scientific and technical fields, including biology, chemistry, economics, geology, mathematics, psychology, meteorology, physics, anthropology, political science, and sociology. The register was created in collaboration

[8] https://aad.archives.gov/aad/series-description.jsp?s=3550. Last retrieved on September 23, 2025.

with several professional organizations, including the American Institute of Biological Sciences, the American Chemical Society, the American Mathematical Society, and the American Psychological Association.

This dataset contains surveys distributed over eight years through various academic societies to respondents who were predominantly academic and research professionals.[9] The content of each record varies slightly by year, but typical entries include details such as name, institution, sex, age, educational background, employment specialty, job function, self-reported income, language ability, citizenship, and memberships in professional organizations. Additional information, such as place of birth (after 1966) and government sponsorship (after 1962), is included in later years. This dataset thus serves as a comprehensive source for understanding the workforce during these periods. The survey response rate was approximately 60% but varied across academic societies. For instance, in 1968, the response rate among biologists was 54%,[10] while around 70% of the eligible individuals in the Register of the American Meteorological Society responded.[11] Additionally, the NSF reported that over 90% of U.S. science doctorates were captured in the 1964 wave of the survey.

This paper uses the 1962–1968 NRSTP data because these four waves include information on the scientists' city of residence. The data was processed by extracting information from the digitized codes, as shown in Figure 1. Subsequently, the codes for each variable are matched with their meaning, which is documented in the photocopies of the codebook films. The raw digitized format consists of thousands of entries, with each line representing an individual record. The values in different positions in each line correspond to different variables (i.e., survey questions). To analyze the data, I first separate these values into their respective variables. Subsequently, I match the numbers with their descriptions based on the codebooks, which are scanned documents without optical character recognition (OCR). I manually clean the codebooks to ensure accurate

[9] 1954, 1958, 1960, 1962, 1964, 1966, 1968, and 1970. The Survey of Doctorate Recipients continues the NRSTP survey after 1970. However, it uses anonymized census data, making it impossible to link scientists to business registration records.

[10] American Institute of Biological Sciences Annual Report 1969.

[11] Bulletin of the American Meteorological Society, Vol. 47, No. 8, August 1966.

mapping between numerical values and descriptions. Where the original scan is faint, certain words are best guesses by ChatGPT-4o based on common nomenclature. ChatGPT excels at this task, as the transformer models are trained to reconstruct incomplete sentences and words.

3.1.2 American Men of Science

I digitized the eleventh edition of AMS which was compiled from 1960 to 1965. First published in 1906 by James McKeen Cattell, the AMS directory is an exceptionally comprehensive source of biographical information for male and female scientists across the United States and Canada. Scholars have used AMS data to investigate how children change the academic productivity of women in science under the setting of baby boom (1946–1964) (Kim and Moser, 2025). I digitized different editions and focused more on the Silent Generation. In addition, some studies have used the *Minerva Jahrbuch der Gelehrten Welt* for academics' information (Iaria, Schwarz, and Waldinger, 2024), it is more of a worldwide directory of academics, yet does not necessarily have the most comprehensive coverage for North America.

The AMS was created through questionnaires with the assistance of various scientific societies, universities, research labs, and an Advisory Committee appointed by the National Academy of Sciences, the National Research Council, and the American Association for the Advancement of Science. As per the Preface to this edition, the criteria for inclusion are:

1. Achievement, through experience and training, of stature in scientific work equivalent to that associated with a doctoral degree, coupled with continued activity in such work.

2. Research activity of high quality in science, evidenced by publication in reputable scientific journals, or, for those whose work cannot be published due to governmental, commercial, or industrial security, by the judgment of peers among immediate co-workers.

3. Attainment of a position of substantial responsibility requiring scientific training and experience equivalent to that described in (1) and (2).

The directory is divided into two sections: Physical and Biological Sciences, and Social and Behavioral Sciences. Only the first section was digitized, because the primary focus of this research is the scientific and technical personnel. This section contains six volumes, and there are around 25,000 entries per volume.

Each entry in the AMS directory provides detailed biographical information about individual scientists, including their education, career history, and areas of research (see the example in Figure 1), providing a comprehensive view of their scientific contributions and professional backgrounds. The records also contain socioeconomic information, which comprises personal data such as the scientist's date of birth, marriage year, number of children, and contact address.

59% of the addresses in the AMS dataset include the zip code, while many addresses only have street names and the city or state. I utilize cloud-based services to enhance the dataset. Specifically, I employ the OpenStreetMap API, which enables the retrieval of the zip code based on the provided addresses. The API helps to increase the proportion of addresses with zip codes from 59% to 64%. This approach not only improves the geographic analysis of scientists, but is also critical for linking scientists across databases (e.g., based on names and zip codes).

3.1.3 Concatenating the Two Data Sources

My NRSTP sample records include 447,317 scientists who responded to the survey between 1962 and 1968. The AMS sample add 59,877 more scientists to the total data sample. 31,468 scientists appear in both datasets based on name and county location. For the overlapping entries, I retain the records in the NRSTP because the variables recorded there are more comprehensive than in AMS. 56% of the AMS scientists appear in the NRSTP records, indicating that the NRSTP has a good record of senior scientists. Thus, AMS serves as a complementary dataset to the NRSTP records on the senior scientists. This results in 475,726 scientists.[12] Because my further analysis requires tangibility of specialty, I drop 17,023 scientists whose work specialties are missing or not correctly recorded. The final dataset contains information on 458,703 scientists.

3.2 Matching Scientists' Data with Other Databases

3.2.1 Matching Scientists and Engineers with Business Registrations

Business registrations in the U.S. are stored by each state's Secretary of State. OpenCorporates gathers this data and distributes it as a single download package.[13] This paper uses data from all jurisdictions (i.e., states) within the U.S. It should be noted that bankruptcies or any other type of litigation against a company are not listed in the records of the Secretary of State. Instead, this type of information would have to be discovered through a litigation search.[14]

The business registry data from OpenCorporates covers 76 million businesses across all U.S. states. The data includes incorporation dates and dissolution dates, as well as the state and registration address for the business. Businesses can be registered in more than one state. For example, a Texas business that also does business in Florida may be registered as a domestic company in Texas and as a foreign company in Florida (Griffin, Kruger, and Mahajan, 2023). In addition, many firms are registered in the state they operate in as well as in Delaware. OpenCorporates covers both and often connects the two registrations through the branch and foreign company variables. The vast majority of businesses formed by the scientists in my sample are domestic firms only.

Although census data, such as the Longitudinal Business Database (LBD), contains business registration information, it only begins in 1976, which is too short a period before the ERISA reform in 1979 to conduct a parallel trend test. OpenCorporates

[12] Scientists whose county location is missing are dropped, because the later matching process relies on both name and location. Zip codes are mapped to counties because people are likely to move or start businesses within a county but not necessarily within the same zip code. The mapping of zip codes to county FIPS codes comes from the U.S. Department of Housing and Urban Development's USPS ZIP Code Crosswalk Files.

[13] I obtained the data under the reference OCESD-14963, data version as of January 2025.

[14] https://www.jonesday.com/en/insights/2012/10/public-disclosure-requirements-for-private-companies-us-vs-europe. Last retrieved on September 23, 2025.

provides business registry data dating back to the 1940s or earlier, depending on the state's records. It also includes the names of the officers linked to the companies, which is essential for matching with the scientists' data. Therefore, OpenCorporates provides the most consistent, publicly available dataset on U.S. business registrations.

The business addresses in OpenCorporates are cleaned using regularization to extract the zip codes and I then match them to the corresponding county. During the period I use to match the addresses with scientists (1945-1990), 57% of the 14,495,168 firms in the dataset possess registered address data. Among these firms, 84% include the zip code. I use OpenStreetMap API to obtain the zip codes for the remaining 16% of the non-standard addresses, thereby obtaining the zip codes for an additional 155,820 addresses and enhancing the coverage of the zip code to 93%.

I use the spaCy library (`en_core_web_lg`) to classify whether an officer's name in the OpenCorporates is likely a human name or a company name. Specifically, the function checks whether the input text includes any entities labeled as "PERSON" by the NLP model. This analysis reveals that 88.08% of the officer names are classified as human names rather than company names, providing insight into the composition of the entities recorded in the dataset.

I then map the OpenCorporates data to the AMS and NRSTP data by name and county FIPS code. I only match scientists to businesses formed between 1945 and 1990 because the scientists in my data sample were born in the 1920s and 1930s. After 1990, they would likely be too old to start a business, and the risk of mistakenly matching scientists with the same name but different identities becomes more significant. In the final data sample,[15] 3.16% of the scientists are found to be associated with at least one business.

3.2.2 Matching Scientists and Engineers with Patent Data

The patent data is from the PatentCity dataset (Bergeaud and Verluise, 2024), which provides the zip code and inventor names of U.S. patents dating back to 1836. Com-

[15] Only scientists with work specialty information are included in the final sample.

pared to the USPTO dataset, which began recording inventor names only in 1976, the PatentCity dataset provides better coverage of historical patent data. It includes records of the "first publication of granted patents," meaning that only the patent applications corresponding to granted patents are included in the dataset.

I then map the patent data with the AMS and NRSTP data by name and county FIPS code. Again, I only match scientists to patents filed between 1900 and 2000 to reduce the risk of mistakenly matching scientists with the same name but different identities. In the final data sample, 9% of the scientists are found to be associated with at least one patent.

3.2.3 Matching Scientists and Engineers with Publication Data

To measure scientific productivity, I match the scientists with their publications and citations from SciSciNet (Lin, Yin, Liu, and Wang, 2023), based on the full data from Microsoft Academic Graph (MAG, now OpenAlex). MAG was updated weekly until December 2021. SciSciNet covers over 134 million scientific publications and millions of external linkages to funding and public uses.

I restrict the data to authors with at least one English-language journal publication between 1900 and 2000. I match the scientists and engineers with the author_ids in MAG, using first and last names, as well as the county FIPS of the author's institutional affiliation. Based on the birth year of the scientists and engineers, I further restrict the matched publications to scientists with no publications after 2005. In the final data sample, 10% of the scientists are found to be associated with at least one published paper.

3.3 Descriptive Statistics

My final sample consists of 458,703 scientists with recorded county FIPS codes and work specialties. This section documents the characteristics of the scientists in my sample.

Gender The AMS dataset lacks gender information, so I supplement it with the gender guesser library. The gender guesser tool utilizes a dataset of approximately 40,000 first names and their associated genders, covering most first names in European countries. For each scientist, I first check the NRSTP for gender information and used it if available. If not, I apply the gender guesser to predict the gender based on the scientist's first name. The sample of scientists and engineers is dominated by males, with 417,903 male scientists and 38,895 female scientists. This is consistent with the literature.

Cohort The NRSTP dataset does not include the date of birth as AMS does, so I develop a method to predict the scientists' year of birth based on the Year of Highest Degree and the Level of Highest Degree recorded in the NRSTP. I assume that scientists typically obtain their PhD (or higher, such as MD) around the age of 30, a Master's degree around the age of 25, and a Bachelor's degree around the age of 22. Using these assumptions, I estimate the year of birth by subtracting the predicted age at the time the highest degree was obtained from the Year of the Highest Degree, thereby improving the overall coverage of missing birth year information. The overall sample is dominated by the Silent Generation (i.e., born between 1928 and 1945). They grew up during the Great Depression and World War II, which shaped a more risk-averse and pragmatic outlook (Figure A3). One may argue that, since the scientists were born in the 1930s–1940s, they may have been too old when the ERISA reform occurred. However, research has shown that the mean age at founding for the fastest-growing startups is 45 (Azoulay, Jones, Kim, and Miranda, 2020).

Education The data sample comprises 458,703 scientists, including 160,082 PhD holders and 10,996 MD holders. The average year in which the scientists obtained their highest degree is 1954. University names are standardized by mapping them to the Integrated Postsecondary Education Data System using both the institution's name and city location. The top three alma maters among the scientists are the University of Michigan-Ann Arbor, Columbia University, and the University of California-Berkeley, while other elite institutions such as Harvard University and MIT are also common

(Table A2).

Geographical Location The majority of scientists are concentrated around San Francisco, Los Angeles, and counties in New England (Figure 2). However, it is worth noting that there are also concentrations of scientists in the central U.S.[16] This concentration suggests that the critical expertise and resources were likely pooled in specific regions, possibly due to the specialized infrastructure or proximity to major research institutions and contractors for government programs (such as the defense program and space program).

Income The scientists earned more than the general population in the lower and middle quantiles (Table A3). Income inequality within the scientific community is less than that of the overall U.S. population. The median wage is higher than the general population, yet the top 1% is lower. These reflect the relatively standardized wage structures within scientific professions.

Employment Most of the scientists are employed in private industry or business, while a significant number also work in colleges and universities (Table A4). The proportion of scientists and engineers in private industry is comparable to that in academia. Within the private sector, the top employers are typically in the chemical manufacturing and petroleum-related industries, the electrical and electronics sectors, and large aerospace and defense contractors (Table A5).

Work Specialty A major challenge was to compile the work specialties into an individual-year panel. The NRSTP generates a sequence of identifiers for each specialty in each wave of the survey. However, these identifiers vary across waves, and the classification of specialties changed year by year. For instance, *Probability and Statistics* was later divided into two separate specialties: *Probability* and *Statistics*. To link

[16]For example, during the Cold War, Natrona County (FIPS 56025) in Wyoming was involved in uranium mining, which was crucial for nuclear weapon development. El Paso County (FIPS 08041) in Colorado is home to the North American Aerospace Defense Command. Additionally, Pima County (FIPS 04019) in Arizona housed a Titan II missile complex, which was operational from 1963 to 1987.

specialties across years, I standardize names and manually merge or split the specialties as needed. If a scientist appears in multiple waves of the NRSTP survey, I retain the most recent first work specialty as their specialty. I also show that scientists typically do not change the tangibility of their specialty (Table A10). The data sample reveals a strong educational background concentration in Chemistry, the Theory and Practice of Computation, and Physics (Table A6).

Patents and Publications The average publication rate among scientists is slightly higher than the patent rate, which in turn exceeds the business formation rate (Table 1). Most businesses founded by scientists do not have a granted patent. On average, each scientist publishes two papers in their lifetime, with a median citation count of 11 and a typical coauthor count of one to two. While most publications are not linked to patents, some highly influential papers are cited by approximately 30,000 patents. There is a weak correlation between business formation activity and both patenting and publishing activity, indicating a limited association between these factors (Table A7). This suggests that scientific output and intellectual property generation do not strongly predict entrepreneurial activity among scientists. Figure A4 shows the number of patents filed per scientist over this time. The data indicate that scientists are most active in patenting during their 30s and 40s. Patenting activity in the sample declines markedly after 1970, and by the time of the ERISA shock in 1979, it had nearly ceased altogether.

4 VC on Scientists' Entrepreneurial Entry

I use the 1979 ERISA reform as an exogenous shock that led to the large-scale emergence of VC as a financial intermediary for two reasons. First, this reform is unique in its significant impact on VC fundraising, as one of the few regulatory changes to have such an effect. While the capital gains tax cut in the 1980s could also have influenced VC investments, most VC investors post-1980 were tax-exempt institutions, such as pension funds, endowments, and trusts, so the supply effect of this tax cut was small

(Gompers, 1994; Gompers and Lerner, 1999). Second, the early-stage equity investment landscape of the 1980s had not yet developed a standardized approach with similar term sheet structures. Equity investment in small businesses was primarily provided by individuals, with little involvement from financial intermediaries. Furthermore, angel investment was not popularized until the 1990s. The ERISA reform played an important role in establishing VC as a key financial intermediary in equity investment. After the ERISA reform, both the number of deals and the total investment amount surged, as illustrated in Figure A2.

I first show that in my data sample, business formation steadily increases over the sample period, with no abrupt change around the 1979 ERISA reform (Figure 4, numbers of business formed are normalized to 1978). Business formation by scientists more than doubled following the ERISA reform, indicating its unique impact on scientists. Notably, the total business formation rate for the scientists from 1945 to 1990 was 3.15%. In contrast, during 1971–1978, the seven years preceding ERISA, the rate was only 0.78%. Although the rate increased to 1.49% in the seven years following ERISA, it remained below the 3% rate observed in the general population.[17] This indicates that scientists have a lower propensity to start businesses than the general population.

4.1 Measuring the tangibility of specialties

The 1979 ERISA reform represents a one-off exogenous shock. As the ERISA reform did not take place until 1979, scientists could not have chosen their work specialties based on anticipated VC funding in the 1960s. Even if their choice of work specialty was influenced by anticipated funding opportunities, they would likely have favored fields more suitable for bank lending. Therefore, scientists' work specialties provide exogenous cross-sectional variation in the exposure to the VC shock, and thus for a Difference-in-Differences (DiD) design.

The cross-sectional variation is based on the assumption that scientists working in fields more reliant on intangible capital likely faced greater exposure to the ERISA

[17]The 3% figure is taken from Wallskog (2025), which is based on census data.

shock. This is consistent with the literature that finds that the scalability of intangibles can enable home-run success (Haskel and Westlake, 2017). Moreover, the assumption is grounded in the fact that banks do not lend to intangible businesses for two reasons. First, asset-based lending depends on the liquidation value of tangible assets that can be pledged as collateral; intangible businesses typically lack such assets. Second, cash flow-based lending relies on stable operational cash flows, which early-stage startups often do not have. Consequently, intangible startups are less likely to receive bank financing.

In this context, I define tangible specialties as those associated with physical products or processes (e.g., a machine or manufacturing method), whereas intangible specialties are related to non-physical outputs, such as software and algorithms.

To distinguish between tangible and intangible work specialties, I utilize a large language model with the word embedding method (Ash and Hansen, 2023). I begin by constructing a sample of publicly listed U.S. firms and collecting their company descriptions from Compustat for the years 1985–1990. I then restrict the sample to firms with two-digit SIC codes between 20–49 and 71–79, excluding sectors like wholesale and retail where scientists are less likely to start businesses. Firms are double-sorted by capital intensity (in descending order) and the share of intangible assets (in ascending order).[18] Relying solely on the share of intangible assets on balance sheets as a proxy for intangible assets is potentially misleading, as many forms of intangible capital (such as know-how and customer capital) are not captured in the book value due to the inherently conservative principles of accounting (Gourio and Rudanko, 2014). The company descriptions of the top 100 firms are used to represent tangible specialties, while those of the bottom 100 firms represent intangible specialties.

With the tangible and intangible corpora, I employ GPT-o3 to construct two distinct dictionaries of scientific specialties, one associated with tangible-intensive firms and the other with intangible-intensive firms. The prompt is described in detail in Section A. Each resulting dictionary comprises 20 specialties that best align with the respective

[18] Capital intensity is defined as the ratio of capital expenditures to total assets. The share of intangible assets is the ratio of intangible assets to total assets.

firm's technology and product. The contents of these dictionaries are listed in Table A8.

I then embed both dictionaries, along with the work specialties, using SciBERT. Word embedding provides a more robust approach than the bag-of-words method for measuring the similarity between a dictionary entry and a word by capturing semantic relationships in a continuous vector space. Unlike the bag-of-words approach, which relies on word frequency and ignores context, embeddings account for meaning and word associations, enabling more accurate comparisons (Jha, Liu, and Manela, 2025; Li, Mai, Shen, and Yan, 2021). This is particularly valuable in my context, as it handles the synonyms of scientific disciplines more effectively. SciBERT is a transformer-based language model specifically trained for scientific text. Developed by Beltagy, Lo, and Cohan (2019), it is based on BERT but pre-trained on a large corpus of scientific literature, including papers from Semantic Scholar. Its domain-specific training allows it to better understand technical terminology and contextual nuances in scientific texts compared to general-purpose language models.

The intangible (tangible) score measures the textual similarity between a scientist's specialty and the intangible (tangible) specialty dictionary. Table 1 shows the descriptive statistics of the scores. The distribution of scores is presented in Figure A6. Some specialties exhibit similarity to both the tangible and intangible dictionaries. For instance, insect toxicology exhibits minimal difference between tangible and intangible similarity scores. This suggests that textual similarity alone does not clearly categorize this specialty as either tangible or intangible. For interpretability in regression analysis, I define a binary variable: a scientist is classified as intangible if the difference between their intangible and tangible scores falls in the top quartile, and tangible if it falls in the bottom quartile. Specialties with intermediate differences remain unclassified. I also include the continuous scores in robustness analyses and find similar results.

For reproducibility, I replicate the dictionary construction procedure using GPT-o4-mini. The resulting dictionaries are reported in Table A9. The similarity between the tangible scores generated by GPT-o3 and GPT-o4-mini is 0.980, while the similarity for intangible scores is 0.988. Given the high concordance across models, I rely on the

GPT-3-based scores for the main analyses.

Table A11 compares the observables of scientists based on their tangible and intangible specialties. The data reveal that female scientists are more likely to have intangible specialties. Scientists with tangible specialties are more frequently associated with government programs in agriculture, atomic energy, and natural resources. In contrast, intangible specialties are more closely linked to government programs related to education, defense, and space.

Table A12 lists the companies with the highest proportion of employees with tangible and intangible specialties. The results indicate that companies operating in computing, data analytics, and systems development exhibit a higher concentration of employees with intangible specialties. Conversely, companies engaged in materials manufacturing and automotive parts employ a greater share of workers specializing in tangible assets.

4.2 Effects of VC on Scientists' Entrepreneurial Entry

The linear probability model with a DiD estimator is:[19]

$$Y_{it} = \alpha_i + \delta_t + \beta Intangible_i * Post1979_t + \text{Controls} + \epsilon_{it} \tag{1}$$

Y_{it} is a binary variable of business formation by scientists i in year t. $Intangible_i$ is a binary variable that equals one if the scientist's work specialty is classified as intangible. $Post1979_t$ is an indicator variable for the post-ERISA reform period. α_i and δ_t denote individual and year fixed effects respectively. Individual fixed effects capture time-invariant determinants of business formation of individual scientists, such as gender and age. Year fixed effects control for aggregate shocks and common trends in business formation activity produced by legal and institutional changes at the federal level, such

[19] I did not use Logit model for two reasons. First, unlike OLS, Logit model relies on maximum likelihood estimation (MLE), which is more sensitive to the distribution of the dependent variable. Given that StartBusiness is highly imbalanced, MLE may struggle to identify significant effects (Timoneda, 2021). Second, in nonlinear models such as Logit, the DiD estimator does not yield a straightforward interpretation as an average treatment effect. The parallel trends assumption does not naturally hold in nonlinear models, and using fixed effects can lead to the exclusion of groups with only 0s or 1s, reducing the sample size and potentially introducing bias. Nevertheless, I include results with non-linear models such as Logit and Poisson in the appendix Table A14.

as the Economic Recovery Tax Act of 1981. Each scientist is retained in the data only until the year when they start a business, such that β can be interpreted as the differential change in the hazard of business formation after ERISA by scientists with intangible and tangible specialties (Basker and Simcoe, 2021). The results are clustered at the individual level, but all results remain significant with county level clustering.

One concern is that scientists shift their occupational specialty when local VC activity expands. For instance, a mechanical engineer might acquire computer science skills and launch a software firm in response to increased VC activity. In this scenario, the VC deal flow becomes a potential omitted variable. Although I do not observe the tangibility of occupational specialties over time, I control for the number of VC deals in each county-year. Table A15 includes other combinations of control variables and county-year fixed effects. The results are significant and robust.

4.2.1 Business Formation

The results in Table 3 show that following the 1979 ERISA deregulation, scientists with more intangible work specialties are significantly more likely to establish new ventures. The results are robust by adding a control in Column (2), year fixed effects in Column (3), and individual scientist fixed effects in Column (4).

Figure 4 plots the coefficients and 95% confidence interval for the year interactions with *Intangible* in Equation 1, using the full scientist-year panel. The beta in each year estimates the probability of business formation by a scientist with an intangible work specialty relative to a scientist with a tangible specialty. The figure shows that the parallel trends assumption is satisfied, indicating that, in the absence of treatment, the treatment and control groups would have followed similar trends over time.

Overall, the results indicate that scientists with intangible work specialties are 0.05 percentage points more likely to start a business than those with tangible work specialties. Given scientists' overall business formation rate of 0.78% before the ERISA shock, this corresponds to a relative effect of approximately 6.41%. This evidence suggests that the influx of private capital effectively alleviates the financial constraints faced by scientists, thereby fostering entrepreneurial and innovation activity.

4.2.2 Robustness Checks

I now consider the robustness of my main results.

Placebo Analysis To validate the DiD design, I conduct a placebo analysis by exploiting the spatial variation in VC presence. If the estimated effect captures the influence of VC, there should be little impact observed in counties without VC activity, as investment opportunities generally circulate within local spaces (Sorenson and Stuart, 2001). Table A16 presents the results separately by VC presence. Panel A restricts the sample to counties with at least one early-stage VC deal during the sample period. Panel B includes scientists residing in counties without any VC presence. The results show a significant effect only in Panel A, while no statistically significant effect is observed in counties without VC presence, consistent with the interpretation that the estimated effects are driven by exposure to VC.

Moreover, one may argue that the credit crunches resulting from Regulation Q[20] coincided with the ERISA reform, and thus the effect may instead be driven by scientists' limited access to bank credit, which pushed them toward VC funding. However, negative credit shocks to young firms can reduce overall business formation. To test this, I use intrastate bank deregulation—an event within the sample period that led to banking industry consolidation—as a source of negative credit supply to young firms, because the consolidation increased banks' bargaining power over small firms (Chava, Oettl, Subramanian, and Subramanian, 2013; Hombert and Matray, 2017). The staggered DiD results in Figure A7 show that the deregulation reduced scientists' business formation. Thus, the positive effect I document is unlikely to be driven by reduced bank credit. Instead, it implies that, although credit crunches may reduce entry, the increased supply of VC more than offsets this decline, leading to a net positive effect on business formation.

[20]Regulation Q prohibited banks from paying interest on demand deposits and imposed caps on savings deposit rates to limit competition for funds.

Measurement of intangibility I test the main results with different definitions and cutoff thresholds of the key independent variable, *Intangible*. Table A17 shows that when I replace the binary definition of Intangible with a continuous variable as the cross-sectional variable for the second difference, the results remain consistent with the binary regression. Scientists with a specialty that has a higher intangibility score are more likely to start a business after the ERISA shock, whereas those with a higher tangibility score are not affected by the shock.

Shocks concurrent with ERISA I address the potential concern that the main effect is primarily driven by the progress of intangible technology, whereby the booming intangible industry in information technology coincided with the ERISA shock. For example, information technology suddenly enabled enormous business opportunities in 1979. As a result, computer scientists and electronic engineers may drive the results. Table A18 presents the results of excluding Silicon Valley-related specialties from the sample and shows that the main estimates remain significant.

Legal structure of business registration I address the concern that many firms are registered in Delaware due to the state's legal infrastructure, so a single state may drive the effect of business formation. Table A20 presents the results of excluding Delaware from the data sample and shows that the main estimates are almost unchanged.

4.3 Entrepreneurial Spawning by Private Scientists

Scientists employed in the private sector and those in academia may differ endogenously in their career incentives and human capital accumulation. Industry scientists gain practical experience through real-world applications, which enhances their entrepreneurial capabilities and increases the likelihood of business formation. Would-be entrepreneurs anticipating financing needs are more likely to start businesses when the supply of capital expands (Samila and Sorenson, 2011). In contrast, university scientists tend to focus on fundamental research and scientific advancements, making them less inclined to pursue commercialization or respond to an increase in the VC supply.

Indeed, Figure 5 and Table 4 indicate that the business formation effect is primarily driven by the private sector, and not by those working in universities or the federal government, or who are self-employed. Column (1) shows that private scientists with intangible specialties had a business formation rate that is relatively 0.18 percentage points higher than that of private scientists with tangible specialties. Given the pre-ERISA business formation rate of 1.22%, this effect represents a 14.8% increase relative to the baseline.

While prior work finds that VC connections facilitate university spinouts (Shane and Stuart, 2002), Column (2) shows that university scientists rarely do so, likely reflecting the innate type of university scientists that makes them self-selected into academia.

To further demonstrate the industry spinout effects, Figure 6 shows the bin scatter of firm-level averages of spinout rates against the share of intangible and tangible employees. In the top panel, the share of intangible employees is positively associated with the spinout rate, whereas the bottom panel shows a modest negative slope for the share of tangible employees. Taken together, the evidence supports the view that intangible human capital within firms is a key driver of entrepreneurial spinout activity.

These results are consistent with the literature on entrepreneurial spawning (Babina and Howell, 2024; Gompers et al., 2005) as entrepreneurial spawning occurs when individuals become entrepreneurs because the large bureaucratic companies for which they work are reluctant to fund their entrepreneurial ideas (Gompers et al., 2005). Employees of large firms thus leverage their experience and expertise to create spinout businesses. A widely cited example is Xerox's Palo Alto Research Center (PARC), which developed groundbreaking technologies like laser printing. Despite its innovations, PARC struggled to gain support for commercialization. The executives resisted moving the company beyond its traditional copier business, and most of the value from Xerox's inventions was captured by employees who left to start companies like Adobe and 3Com.

Although engineers employed in large firms may be motivated to leverage their expertise to transfer technology through business formation, the lack of financing for potential startups to commercialize products can hinder entrepreneurial spawning.[21] Moreover, though possessing technical knowledge, engineers may lack the business acumen and network essential for entrepreneurship.[22] The results in this section show that VC reduces the financial constraints and incentivizes private scientists to spinout from their employer.

For a sanity check, I further explore the effect in California counties and non-California counties. Table A19 shows that the size of the spinout effect is almost nine times larger for scientists living in California compared to those living outside of California. This confirms that California was the center of the VC industry and had the most vibrant entrepreneurial ecosystem at that time.

4.4 Quality of Startups

The previous sections show that scientists working in the private sector, instead of universities, were the most responsive to the VC shock. A natural question that arises is whether these scientists represent marginal entrants into entrepreneurship—that is, scientists who were previously unable to obtain funding due to the lower quality of their ideas and who only entered the market following the expansion in the VC supply.

This section relies on an ex ante measure of startup quality, as ex post performance measures are unavailable. I use business registration data to capture the business formation of scientists, so I do not have access to follow-on firm performance data, because these firms are private. While employment and sales data for private firms can be accessed through census data, I do not have such access. Therefore, I use scientists' characteristics ex ante as proxies for the quality of the businesses they started.

Productivity I use self-reported annual gross income in NRSTP as a proxy of productivity. Panel A of Table 5 presents estimates for scientists employed in private industry.

[21] For instance, the companies in the Central Florida Research Park (CFRP) in Orlando have struggled to grow their size and customer base. As a result, the success of the CFRP is still overly tied to the military budget.

[22] As VC funding was pouring into startups that focused not on rockets but on corporate computers, Silicon Valley's engineers were far less dependent on space contracts by 1969 (Miller, 2022).

Columns (1)–(4) stratify the sample by annual gross-income quartiles (Q1–Q4). The coefficient of interaction term is positive and statistically significant in every quartile, rising from 0.0841 in Q1 to 0.2561 in Q4. The monotonic increase implies that scientists with higher productivity responded more strongly to the VC shock. Figure 7 corroborates this result. The estimated VC effect for private-sector scientists increases monotonically across income quartiles, and event study plots confirm that pre-treatment trends are parallel. Columns (5)–(8) repeat the exercise for university scientists. All VC interaction terms are statistically insignificant and display no systematic trend across income quartiles.

Innovation Activity I examine whether the scientists' innovation activity (i.e., patents and publications) is related to business formation after the VC shock. It is worth noting that although patents can serve as loan collateral and so reduce scientists' financial constraints, only about 2.5% of the patents issued in 1980 were pledged within five years (Mann, 2018).[23] Although some studies have shown that one-third of startups used patents to collateralize and raise venture debt, the sample is already limited to VC-backed startups; if the denominator were instead the universe of tech startups, the ratio would be much lower (Hochberg et al., 2018; Serrano and Ziedonis, 2025). Moreover, the scientists in my sample are the inventors rather than the assignees of the patents, so they do not necessarily possess the patents granted for their inventions. Therefore, being an inventor on patents is only used to proxy innovation activity, not financial constraints, in this study.

In Table 6, Panel A, Columns (1) and (2) show that scientists who filed at least one patent and were employed in the private sector were significantly more likely to spin out. The effect is approximately 0.74 percentage points, indicating a 25.5% increase in the likelihood of starting a business with a baseline rate of 2.90% before ERISA for the private scientists who filed at least one patent. This substantial effect aligns with the argument that inventors seek to appropriate the value from their inventions, but

[23]A key deterrent is legal uncertainty. Federal statutes such as the Patent Act treat a security interest as a conditional transfer of title to the creditor, whereas the UCC allows the debtor to retain ownership and merely grants the lender a security interest (Baldwin, 1994).

large firms often capture most of the benefits, creating an incentive for them to spin out. Columns (3) and (4) show that scientists who had published a critical journal article were also more likely to start a business, though the effect is smaller compared to patenting. This suggests that publishing scientific articles is less directly related to commercialization, whereas patenting is more strongly associated with business formation.

Although university scientists were not responsive to the VC shock in general, as demonstrated in Section 4.3, almost half of the scientists in my sample worked in universities, so it is worth examining whether university scientists' patenting and publishing activities were more attractive to VC. In Table 6, Panel B, Columns (5) and (6) show that university scientists who had filed at least one patent were significantly more likely to start a business. However, the share of university scientists who filed patents is low. There is no significant effect of business formation among university scientists who published journal articles, as shown in Columns (7) and (8). This differs from the results in Panel A, which indicate that university scientists were less likely to start a business compared to industry scientists and that publishing papers did not make business formation more likely. Instead, filing patents appears to be a good signal to VC.

Table A21 tests coefficient equality by including indicator variables for whether the scientist was a patent inventor and whether they were a paper author, each interacted with the intangible specialty measure. The patent–intangible interaction has a larger effect size and is statistically significant. The publication–intangible interaction is roughly half as large and is not significant, suggesting that patenting, rather than publishing, was the primary driver of the VC effect on scientists.

A potential concern is that scientists who were inventors or journal article authors may differ systematically from those who were not. The sample is also highly unbalanced, as the majority of scientists had neither been named as inventors on patents nor published journal articles prior to 1979. To address this, I implement propensity score matching on observable characteristics. Table A22 reports results with propensity score matching, which are consistent with those in Table 6.

This section shows that the increased availability of VC incentivized private scientists, especially those with higher productivity and innovation activity, to start a business. This suggests that VC encouraged high-quality scientists who could not have secured funding before to enter entrepreneurship, as opposed to marginal scientists.

5 A Simple Occupational Choice Model

The empirical evidence presented suggests that VC incentivized scientists with intangible specialties to start businesses. This raises two related questions: why are scientists with higher productivity more responsive to the VC; and how much does VC relax the financial constraints of scientists?

The model provides a framework to rationalize the empirical results and quantify the relaxation of financial constraints. I now present a theoretical framework that builds on the financial constraints and selection into entrepreneurship of Evans and Jovanovic (1989) and differentiates physical and intangible capital. The model rationalizes the empirical results that more productive scientists are more responsive to the VC shock. Furthermore, I calibrate the model to quantify the changes in financial constraints before and after ERISA.

5.1 Model Setup

I start the model with two types of scientists: those with intangible specialties and those with tangible specialties. They have homogeneous wealth distribution, denoted by a, which is interpreted as net family assets. They also have homogeneous productivity distribution, denoted by z, which determines the productivity of capital in entrepreneurship. I assume that $\log(z) \sim \mathcal{N}(\mu_z, \sigma_z^2)$, as innovations are usually characterized by a skewed distribution in the literature.[24]

Output is produced with a Cobb-Douglas aggregator over two types of capital: k_1 is

[24] In a similar vein, variables such as deal size and valuation in the VC literature are also skewed and heavy-tailed, so are typically modeled as Pareto or lognormal.

physical capital and k_2 is intangible capital (Crouzet and Eberly, 2023). The parameter η denotes the share of intangible capital in production. Scientists with intangible specialties have a higher $\eta_{\text{intangible}}$, whereas those with tangible specialties have a lower η_{tangible}. The parameter α is the returns-to-scale parameter, which differs between scientists with intangible and tangible specialties.

$$y = z\left(k_1^{1-\eta}k_2^{\eta}\right)^{\alpha}$$

Scientists also face a wage opportunity w, representing the income they would earn if they remained in salaried employment. The wages of scientists with different types of specialties are slightly different based on my data, with $w_{\text{intangible}} > w_{\text{tangible}}$. The occupational choice is static and discrete: a scientist chooses entrepreneurship if the entrepreneurial income exceeds the wage alternative. The entrepreneurial income is given by

$$\max\ \pi(z,k_1,k_2,a) = z\left(k_1^{1-\eta}k_2^{\eta}\right)^{\alpha} - r(k_1+k_2-a), \text{with } 0<\alpha<1,\ r>1,$$

where α is the elasticity of output with respect to capital and r is the gross interest rate. The first term captures revenue as a Cobb-Douglas function of capital and productivity; the second term reflects the cost of financing in excess of own wealth and invested capital.

Due to financing constraints, scientists can invest at most $I = \lambda a + b$ in entrepreneurial activities, where λ captures the access to external capital per unit of wealth. The emergence of VC could relax constraints by providing external financing and so increase λ. One possible channel is through better monitoring.[25]

Only k_1 is pledgeable to reflect the sunkness (i.e., low liquidation value) of intangible

[25]For example, assume the cost of defaulting is nI, where n is the monitoring intensity. The cost of repayment is $r(I-a)$. Therefore, the maximum amount that a person could obtain from external financing is $\lambda = \frac{r}{r-n}a$. If VC increases the monitoring intensity n, then λ increases.

assets imposes tighter liquidity constraints. Borrowing b must satisfy

$$b \leq \phi k_1, \qquad \phi \in [0,1].$$

5.2 Model Results and Implications

This section shows why scientists with higher productivity can be more constrained.

The scientist chooses entrepreneurship if $\pi(z, \lambda, a) \geq w + ra$.

Lemma 1. *The optimal level of investment is*

$$I = \min\left\{ \left(\frac{\alpha \kappa z}{r}\right)^{\frac{1}{1-\alpha}}, \frac{\lambda a}{1-\phi(1-\eta)} \right\}.$$

where the first term denotes the unconstrained optimal investment and the second term the maximum feasible level under financing constraints.

Proof in Appendix B. The higher the productivity z, the higher the unconstrained optimal investment, making it more likely that the entrepreneur is constrained. Venture capital raises λ, which increases the maximum feasible investment level and relaxes financing constraints, particularly for scientists with higher z.

Lemma 1 rationalizes the empirical finding that scientists who are named inventors on patents (i.e., those with higher z) are more responsive to the VC shock. This is because they have a higher optimal level of investment and are therefore more likely to be constrained.

5.3 Model Calibration

This section is to quantify how much VC relaxes financial constraints based on the empirical results. The central parameter in this model is λ, which determines the financial constraints. A higher λ value imposes higher financial constraints and reduces the entrepreneurial entry decision.

The calibration is to find λs that match the entry rate of scientists pre and post ERISA, respectively. The parameters are summarized in Table 7. Whenever available, I use values from the literature. The capital output elasticity α defines the shape of the entrepreneurial production function and the diminishing returns to capital and is recovered from the economic rents accruing to the firm parameter in Crouzet and Eberly (2023). Scientists with intangible and tangible specialites have different α, respectively. The gross interest rate r equals 1.108, which is from the 10-year treasury yield in January 1980. w wage is the gross annual income as self-reported by the scientists in the NRSTP. Wages of scientists who were already self-employed are not included. The average annual gross income is reported around 1965 and scaled to 1980 values in the model.

With the parameters, pre and post ERISA λ are recovered from the observed entry rates before and after the regulatory shock. Figure 9 plots the model-predicted entry probability against wealth for the calibrated $\hat{\lambda}$s, using scientists with intangible specialties and tangible specialties, respectively. The two lines represent the relationship with calibrated $\hat{\lambda}$s before and after the ERISA shock. It shows that ERISA relaxed the financial constraints of intangible scientists by 69.2% through the increase of λ, while only 41.5% for tangible scientists. This pattern is consistent with the data that both tangible and intangible scientists are responsive to the ERISA shock but intangible scientists are even more responsive.

One possible channel inferred from the model is that because tangible scientists have more physical capital to pledge, they need less VC financing and so smaller λs. Subsequently, when the ERISA increased the supply of VC, their changes of λs are smaller than intangible scientists.

Although $\hat{\lambda}$ is an informative measure of the fraction of wealth that can be externally financed, it should be interpreted with caution. The estimate rests on assumptions about the distribution of productivity, represents VC solely through borrowing limits against wealth, and ignores possible interactions between wage and ability. Subject to these caveats, the results highlight the potency of VC in easing financial frictions: despite the market's small size in 1980, the reform substantially lowered scientists' entry barriers. This suggests that regions such as Europe could unlock considerable entrepreneurial potential if pension fund regulations on VC investment were similarly relaxed.

6 Aggregate Effects of VC on Industry Growth

The previous results have shown the causal effect of a change in the supply of VC on scientists' entrepreneurial entry. These findings raise an important question: how do such changes ultimately affect industry productivity and real outcomes? By documenting that VC has a positive effect on industry size and value added, I provide evidence for a channel for the rise of the intangible economy in the U.S.

Consistent with the method used in Section 4.1, I restrict the sample industries to two-digit SIC codes between 20–49 and 71–79. Industries are double-sorted by the average capital intensity (in descending order) and the share of intangible assets (in ascending order) of the firms. The top half of industries are classified as tangible industries, and the bottom half as intangible industries.

I use County Business Patterns (CBP) data files from 1974 to 1984 for the business formation data. The data collection process heavily relied on administrative records, particularly from the Internal Revenue Service (IRS), and existing Census Bureau surveys, with employer-reported information forming the foundation of the CBP. The IRS's quarterly payroll file[26] served as the cornerstone for collecting payroll data, especially for single-establishment employers. The VC investment data comes from VentureXpert. This is the only database that covers the VC and PE deals in 1970s, making it a valuable resource for the analysis in this study (Kaplan and Lerner, 2017).

I estimate the following specification:

$$Y_{ict} = \beta Intangible_i \times Post1979_t + \eta_i + \eta_{ct} + \epsilon_{ict} \tag{2}$$

Y_{ict} represents the outcome variables, including the number of establishments and employment in industry i, county c, and year t. $Intangible_i$ is a binary variable that equals one if the SIC two-digit industry is classified as intangible. $Post1979_t$ is an indicator variable for the post-ERISA reform period. η_i and η_{ct} denote individual and county-year fixed effects respectively. Industry fixed effects capture time-invariant deter-

[26] Treasury Form 941.

minants of industries, such as technological intensity, capital structure, and regulatory environment. County-year fixed effects control for localized macroeconomic shocks, labor market conditions, and demographic trends that vary over time.

Counties with at least one VC investment during 1979–1986 are classified as having VC presence. I do not impose a higher threshold (e.g., more than ten deals), since VentureXpert may not fully capture all VC investments in that period, and a stricter cutoff would exclude many counties that may have had deals. Moreover, VC activity is highly concentrated in the U.S., with only a small fraction of counties having VC presence. Therefore, I adopt a one-deal threshold to allow for a more balanced comparison.

Table 8 presents the effects of VC on industry-level outcomes. Column (2) indicates that, following the ERISA reform, the average number of employees in intangible industries increased by 106. Relative to the pre-ERISA mean of 398 employees per industry, this corresponds to a 27% increase. Columns (3) and (4) demonstrate that this effect is concentrated in counties with VC activity. In these counties, employment in intangible industries increased by an average of 592 post-ERISA. These patterns are consistent across measures based on the number of establishments, both overall and those with 1–4 employees. These findings suggest that VC played a significant role in expanding the size of intangible industries in the post-ERISA period.

A potential concern is that the estimates may simply capture a general rising trend in the intangible industry. To address this issue, I redefine the dependent variable as the industry-level growth rate in each county. Table A23 shows that VC also facilitated the growth rate of intangible industries, not merely their levels.

To assess the robustness of the measure of intangibility, I cross-validate it using the liquidation recovery rate of Property, Plant, and Equipment (PP&E) from Kermani and Ma (2023). The liquidation recovery rates in the data are calculated as the ratio of liquidation value to replacement cost. Industries with lower PPE recovery rates tend to face tighter borrowing constraints and exhibit higher levels of intangibility. Intangibles are especially pronounced in industries where physical assets are more specific. Table A24 examines VC's heterogeneous effects across industries with varying recovery rates.

The results are consistent with those in Table 8, showing a relative decline in industry size post-ERISA among industries with high recovery rates.

These results support the hypothesis that VC contributed to the growth of intangible industries. The findings on the number of establishments are consistent with micro-level evidence of increased entrepreneurial entry by scientists and imply a broader effect on the population outside of scientists.

7 Conclusion

This paper contributes to the growing literature on technology entrepreneurship by demonstrating that an expanded supply of VC can incentivize scientists to create businesses. Exploiting the 1979 ERISA shock to the VC supply, I show that the rate of business formation among scientists doubled, and was especially significant among scientists with intangible specialties. This is because VC alleviates the financing constraints for scientists whose projects lack collateral and in which traditional banks are typically unwilling to invest.

I further investigate the heterogeneity underlying the main results. I find that scientists employed in the private sector exhibited a greater responsiveness to the VC supply shock compared to those affiliated with universities. Private scientists with a higher annual gross income and with prior patenting activity were more likely to spin out. These findings suggest that VC encourages high-quality startups. Through county-industry-level analysis, I demonstrate that these effects ultimately impact real outcomes. Intangible industries in the counties with a VC presence grow more in terms of firm counts and number of employees. To rationalize these empirical findings, I use a simple framework in which scientists decide between remaining wage earners and becoming entrepreneurs. I then calibrate the financial constraints and quantify how VC alleviates them.

This paper has clear policy implications. Governments are increasingly turning to policies to improve access to financing for private firms. VC plays an important role as a specialized financial intermediary in incentivizing technology entrepreneurship, yet the policy tools usually center around publicly-backed VC and tax incentives for equity investors. The results in this paper show that the allocation of pension funds to VCs can incentivize high-quality technology entrepreneurship. This resonates with the ongoing debate over increasing pension fund allocations to VC in the United Kingdom[27]; and a recent White House Executive Order directs the DOL to clarify ERISA guidance so that individual 401(k) investors can access alternative assets (e.g., private equity), expanding access beyond large pension fund managers[28]. By exploiting the 1979 ERISA reform, which permitted private pension funds to invest in VC, I show that such policy changes can produce substantial spillover effects, prompting scientists to establish new ventures and catalyze innovation-led growth.

[27]https://www.gov.uk/government/collections/mansion-house-2023. Last retrieved on September 23, 2025.

[28]https://www.whitehouse.gov/presidential-actions/2025/08/democratizing-access-to-alternative-assets-for-401k-investors/. Last retrieved on September 23, 2025.

References

Acemoglu, Daron, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr, 2018, Innovation, Reallocation, and Growth, *American Economic Review* 108, 3491.

Aghion, Philippe, and Peter Howitt, 1992, A Model of Growth Through Creative Destruction, *Econometrica* 60, 323–351.

Akcigit, Ufuk, and William R. Kerr, 2018, Growth through Heterogeneous Innovations, *Journal of Political Economy* 126, 1374–1443.

Ash, Elliott, and Stephen Hansen, 2023, Text Algorithms in Economics, *Annual Review of Economics* 15, 659–688.

Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda, 2020, Age and High-Growth Entrepreneurship, *American Economic Review: Insights* 2, 65–82.

Babina, Tania, Asaf Bernstein, and Filippo Mezzanotti, 2023, Financial Disruptions and the Organization of Innovation: Evidence from the Great Depression, *Review of Financial Studies* 36, 4271–4317.

Babina, Tania, and Sabrina T. Howell, 2024, Entrepreneurial Spillovers from Corporate R&D, *Journal of Labor Economics* 42, 469–509.

Baldwin, Shawn K., 1994, To Promote the Progress of Science and Useful Arts: A Role for Federal Regulation of Intellectual Property As Collateral, *University of Pennsylvania Law Review* 143, 1701–1738.

Basker, Emek, and Timothy Simcoe, 2021, Upstream, Downstream: Diffusion and Impacts of the Universal Product Code, *Journal of Political Economy* 129, 1252–1286.

Beck, Thorsten, Robin Döttling, Thomas Lambert, and Mathijs van Dijk, 2023, Liquidity Creation, Investment, and Growth, *Journal of Economic Growth* 28, 297–336.

Beltagy, Iz, Kyle Lo, and Arman Cohan, 2019, SciBERT: A Pretrained Language Model for Scientific Text, *arXiv preprint* 1903.10676.

Benmelech, Efraim, Nitish Kumar, and Raghuram Rajan, 2022, The Secured Credit Premium and the Issuance of Secured Debt, *Journal of Financial Economics* 146, 143–171.

Benmelech, Efraim, Nitish Kumar, and Raghuram Rajan, 2024, The Decline of Secured Debt, *Journal of Finance* 79, 35–93.

Bergeaud, Antonin, and Cyril Verluise, 2024, A New Dataset to Study a Century of Innovation in Europe and in the US, *Research Policy* 53, 104903.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend, 2016, The Impact of Venture Capital Monitoring, *Journal of Finance* 71, 1591–1622.

Callaway, Brantly, and Pedro H.C. Sant'Anna, 2021, Difference-in-Differences with Multiple Time Periods, *Journal of Econometrics* 225, 200–230.

Chava, Sudheer, Alexander Oettl, Ajay Subramanian, and Krishnamurthy V. Subramanian, 2013, Banking deregulation and innovation, *Journal of Financial Economics* 109, 759–774.

Chen, Jun, and Michael Ewens, 2025, Venture Capital and Startup Agglomeration, *Journal of Finance* 80, 2153–2198.

Christensen, Clayton M., 2011, *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail* (Harvard Business Review Press).

Corradin, Stefano, and Alexander Popov, 2015, House Prices, Home Equity Borrowing, and Entrepreneurship, *Review of Financial Studies* 28, 2399–2428.

Corrado, Carol, Charles Hulten, and Daniel Sichel, 2009, Intangible Capital and U.S. Economic Growth, *Review of Income and Wealth* 55, 661–685.

Crouzet, Nicolas, and Janice Eberly, 2021, Intangibles, Markups, and the Measurement of Productivity Growth, *Journal of Monetary Economics* 124, S92–S109.

Crouzet, Nicolas, and Janice Eberly, 2023, Rents and Intangible Capital: A Q+ Framework, *Journal of Finance* 78, 339–360.

Crouzet, Nicolas, Janice Eberly, Andrea L. Eisfeldt, and Dimitris Papanikolaou, 2022, Intangible Capital, Non-Rivalry, and Growth, *SSRN Electronic Journal* .

Cunningham, Colleen, Florian Ederer, and Song Ma, 2021, Killer Acquisitions, *Journal of Political Economy* 129, 649–702.

Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters, 2020, A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790, *NBER Working Paper* .

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2013, Organization Capital and the Cross-Section of Expected Returns, *Journal of Finance* 68, 1365–1406.

Eisfeldt, Andrea L., and Dimitris Papanikolaou, 2014, The Value and Ownership of Intangible Capital, *American Economic Review* 104, 189–94.

Evans, David S, and Boyan Jovanovic, 1989, An Estimated Model of Entrepreneurial Choice under Liquidity Constraints, *Journal of Political Economy* 97, 808–827.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of Experimentation and the Evolution of Venture Capital, *Journal of Financial Economics* 128, 422–442.

Fehder, Daniel C., Naomi Hausman, and Yael V. Hochberg, 2025, Innovation and Capital, *Journal of Financial Economics* 169, 104029.

Florida, Richard L., and Martin Kenney, 1988, Venture Capital-Financed Innovation and Technological Change in the USA, *Research Policy* 17, 119–137.

Gompers, Paul, Josh Lerner, and David Scharfstein, 2005, Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999, *Journal of Finance* 60, 577–614.

Gompers, Paul A., 1994, The Rise and Fall of Venture Capital, *Business and Economic History* 23, 1–26.

Gompers, Paul A., 1995, Optimal Investment, Monitoring, and the Staging of Venture Capital, *Journal of Finance* 50, 285–313.

Gompers, Paul A., and Josh Lerner, 1999, What Drives Venture Capital Fundraising?

González-Uribe, Juanita, 2020, Exchanges of Innovation Resources Inside Venture Capital Portfolios, *Journal of Financial Economics* 135, 144–168.

Gornall, Will, and Ilya A. Strebulaev, 2021, The Economic Impact of Venture Capital: Evidence from Public Companies, *SSRN Electronic Journal* .

Gourio, François, and Leena Rudanko, 2014, Can Intangible Capital Explain Cyclical Movements in the Labor Wedge?, *American Economic Review* 104, 183–88.

Griffin, John M., Samuel Kruger, and Prateek Mahajan, 2023, Did FinTech Lenders Facilitate PPP Fraud?, *Journal of Finance* 78, 1777–1827.

Haskel, Jonathan, and Stian Westlake, 2017, *Capitalism without Capital: The Rise of the Intangible Economy* (Princeton University Press).

Hellmann, Thomas, Laura Lindsey, and Manju Puri, 2008, Building Relationships Early: Banks in Venture Capital, *Review of Financial Studies* 21, 513–541.

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.

Hombert, Johan, and Adrien Matray, 2017, The Real Effects of Lending Relationships on Innovative Firms and Inventor Mobility, *Review of Financial Studies* 30, 2413–2445.

Howell, Sabrina T., 2017, Financing Innovation: Evidence from R&D Grants, *American Economic Review* 107, 1136–1164.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger, 2024, Gender Gaps in Academia: Global Evidence Over the Twentieth Century, *SSRN Electronic Journal* .

Jayaratne, Jith, and Philip E. Strahan, 1996, The Finance-Growth Nexus: Evidence from Bank Branch Deregulation, *Quarterly Journal of Economics* 111, 639–670.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2025, Does Finance Benefit Society? A Language Embedding Approach, *Review of Financial Studies* hhaf012.

Kaplan, Steven N, and Josh Lerner, 2017, Venture Capital Data: Opportunities and Challenges, in John Haltiwanger, Erik Hurst, Javier Miranda, and Antoinette Schoar, eds., *Measuring Entrepreneurial Businesses: Current Knowledge and Challenges*, 413–431 (University of Chicago Press, Chicago).

Kermani, Amir, and Yueran Ma, 2023, Asset Specificity of Nonfinancial Firms, *Quarterly Journal of Economics* 138, 205–264.

Kerr, Sari Pekkala, William R. Kerr, and Ramana Nanda, 2022, House Prices, Home Equity and Entrepreneurship: Evidence from U.S. Census Micro Data, *Journal of Monetary Economics* 130, 103–119.

Kerr, William R., and Ramana Nanda, 2009, Democratizing Entry: Banking Deregulations, Financing Constraints, and Entrepreneurship, *Journal of Financial Economics* 94, 124–149.

Kim, Scott, and Petra Moser, 2025, Women in Science. Lessons From the Baby Boom, *Econometrica* 93, 1521–1560.

Korteweg, Arthur, and Berk Sensoy, 2023, How Unique is VC's American History?, *Journal of Economic Literature* 61, 274–294.

Kortum, Samuel, and Josh Lerner, 2000, Assessing the Contribution of Venture Capital to Innovation, *The RAND Journal of Economics* 31, 674–692.

Lerner, Josh, and Ramana Nanda, 2020, Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn, *Journal of Economic Perspectives* 34, 237–261.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2021, Measuring Corporate Culture Using Machine Learning, *Review of Financial Studies* 34, 3265–3315.

Lin, Zihang, Yian Yin, Lu Liu, and Dashun Wang, 2023, SciSciNet: A Large-Scale Open Data Lake for the Science of Science Research, *Scientific Data* 10, 1–22.

Ma, Song, 2025, Technological Obsolescence, *Review of Financial Studies* .

Mann, William, 2018, Creditor Rights and Innovation: Evidence from Patent Collateral, *Journal of Financial Economics* 130, 25–47.

Miller, Christopher, 2022, *Chip War: The Fight for the World's Most Critical Technology* (Scribner).

Morse, Adair, 2024, Venture Debt, in Douglas Cumming, and Benjamin Hammer, eds., *The Palgrave Encyclopedia of Private Equity* (Palgrave Macmillan, Cham).

Nanda, Ramana, and Tom Nicholas, 2014, Did Bank Distress Stifle Innovation During the Great Depression?, *Journal of Financial Economics* 114, 273–292.

Nanda, Ramana, and Matthew Rhodes-Kropf, 2017, Financing Risk and Innovation, *Management Science* 63, 901–918.

Nicholas, Tom, 2019, *VC: An American History* (Harvard University Press).

Peters, Ryan H., and Lucian A. Taylor, 2017, Intangible Capital and the Investment-Q Relation, *Journal of Financial Economics* 123, 251–272.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman, 2018, Distributional National Accounts: Methods and Estimates for the United States, *Quarterly Journal of Economics* 133, 553–609.

Robb, Alicia M., and David T. Robinson, 2014, The Capital Structure Decisions of New Firms, *Review of Financial Studies* 27, 153–179.

Samila, Sampsa, and Olav Sorenson, 2011, Venture Capital, Entrepreneurship, and Economic Growth, *Review of Economics and Statistics* 93, 338–349.

Schumpeter, Joseph Alois, 1983, *The Theory of Economic Development: An Inquiry Into Profits, Credit, Interest, and the Business Cycle*, number 1 (Transaction Publishers).

Serrano, Carlos J., and Rosemarie H. Ziedonis, 2025, Do Patent Assets Have a Second Life When Startups Fail? An Analysis of the Redeployment Likelihood and Mode of Transfer, *Strategic Management Journal* 46, 82–113.

Shane, Scott, and Toby E. Stuart, 2002, Organizational Endowments and the Performance of University Start-ups, *Management Science* 48, 154–170.

Sorenson, Olav, and Toby E. Stuart, 2001, Syndication Networks and the Spatial Distribution of Venture Capital Investments, *American Journal of Sociology* 10, 1546–1588.

Sun, Liyang, and Sarah Abraham, 2021, Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects, *Journal of Econometrics* 225, 175–199.

Timoneda, Joan C., 2021, Estimating Group Fixed Effects in Panel Data with a Binary Dependent Variable: How the LPM Outperforms Logistic Regression in Rare Events Data, *Social Science Research* 93, 102486.

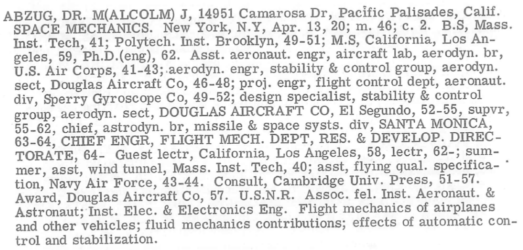
Wallskog, Melanie, 2025, Entrepreneurial Spillovers across Coworkers, *Journal of Financial Economics* 170, 104077.

Zock, Richard, 1980, *Small Business Access to Capital Markets Through Pension Funds* (Interagency Task Force on Small Business Finance, Washington, D.C.).

Figures

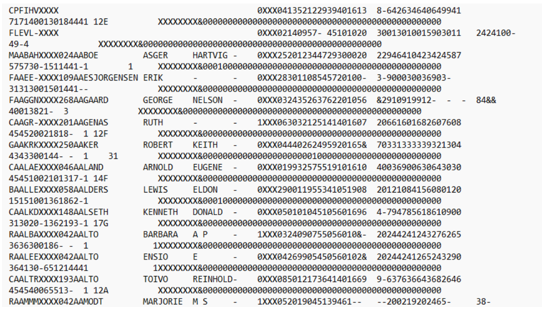
Figure 1: Examples of the Raw Data

(**a**) AWS

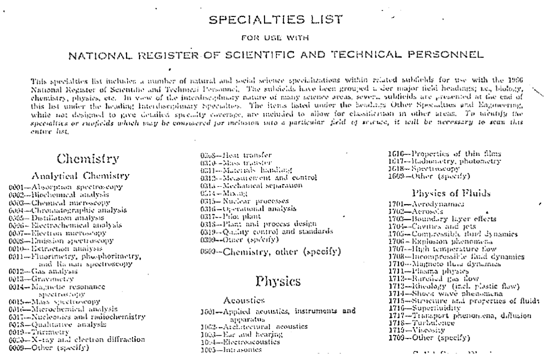


ABZUG, DR. M(ALCOLM) J, 14951 Camarosa Dr, Pacific Palisades, Calif. SPACE MECHANICS. New York, N.Y, Apr. 13, 20; m. 46; c. 2. B.S, Mass. Inst. Tech, 41; Polytech. Inst. Brooklyn, 49-51; M.S, California, Los Angeles, 59, Ph.D.(eng), 62. Asst. aeronaut. engr, aircraft lab, aerodyn. br, U.S. Air Corps, 41-43; aerodyn. engr, stability & control group, aerodyn. sect, Douglas Aircraft Co, 46-48; proj. engr, flight control dept, aeronaut. div, Sperry Gyroscope Co, 49-52; design specialist, stability & control group, aerodyn. sect, DOUGLAS AIRCRAFT CO, El Segundo, 52-55, supvr, 55-62, chief, astrodyn. br, missile & space systs. div, SANTA MONICA, 63-64, CHIEF ENGR, FLIGHT MECH. DEPT, RES. & DEVELOP. DIRECTORATE, 64- Guest lectr, California, Los Angeles, 58, lectr, 62-; summer, asst, wind tunnel, Mass. Inst. Tech, 40; asst, flying qual. specification, Navy Air Force, 43-44. Consult, Cambridge Univ. Press, 51-57. Award, Douglas Aircraft Co, 57. U.S.N.R. Assoc. fel. Inst. Aeronaut. & Astronaut; Inst. Elec. & Electronics Eng. Flight mechanics of airplanes and other vehicles; fluid mechanics contributions; effects of automatic control and stabilization.

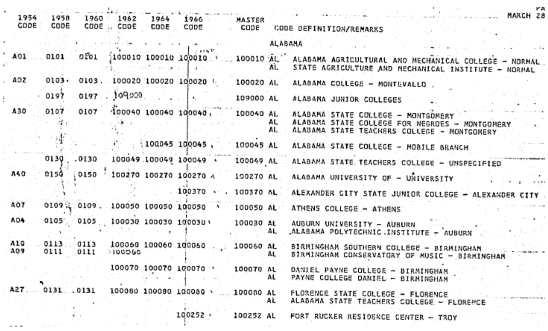
(**b**) NRSTP: Entries



(**c**) NRSTP: Work Specialty Codebook

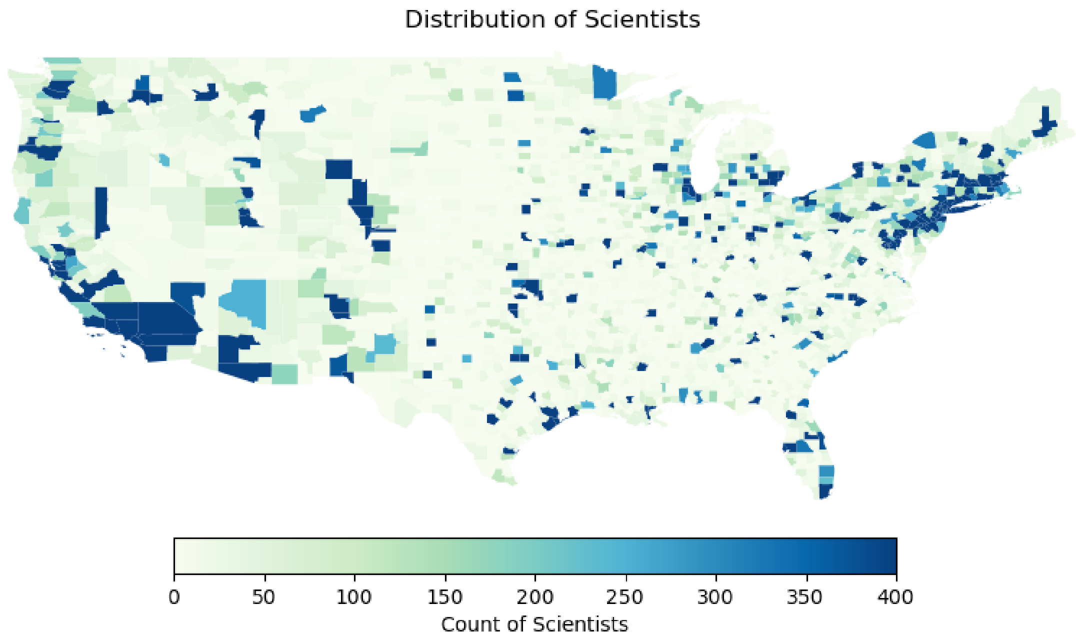


(**d**) NRSTP: University Codebook



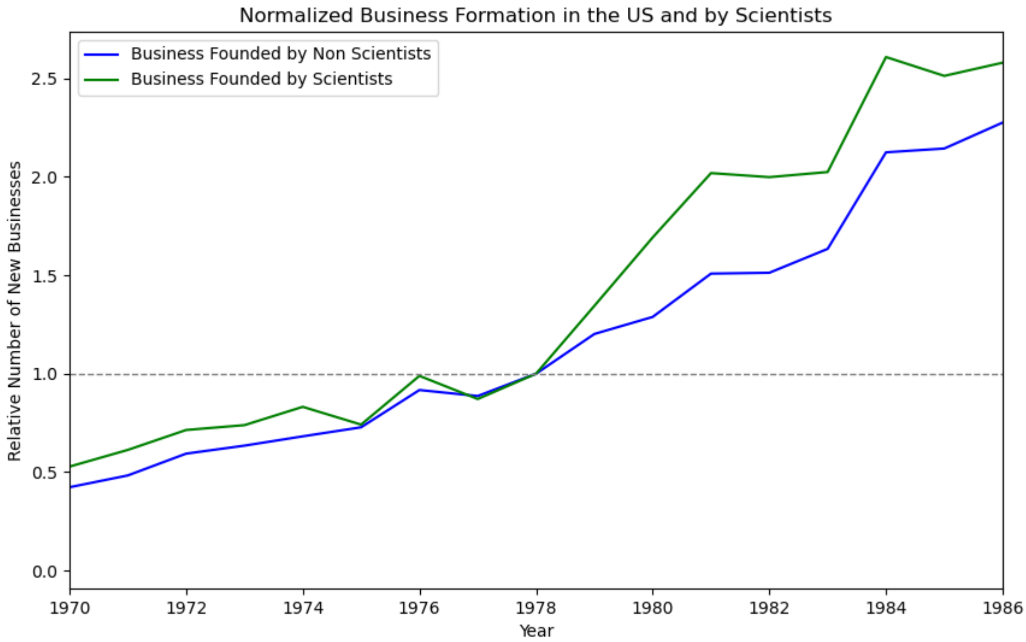
Notes: (a) is an example of an AMS entry. Dr. Malcolm J. Abzug (April 13, 1920), an expert in space and flight mechanics, held prominent roles in aerodynamics, missile systems, and space research. Educated at MIT (Bachelor), Polytechnic Institute of Brooklyn (Master's), and UCLA (PhD), he contributed significantly to Douglas Aircraft Co. and U.S. Air Corps. His research focused on flight mechanics, fluid mechanics, and control systems. (b) shows the raw dataset from the NRSTP. Each line represents one scientist's entry. The dataset is structured so that different positions within a row correspond to different variables. Each variable is encoded using specific numerical or categorical codes, where the position of the code determines which variable it represents. (c) and (d) display the original codebooks of the NRSTP. These codebooks serve as reference documents that map each code in the dataset to its corresponding meaning. When the ORC could not accurately identify certain words, a large language model was used to fill in missing or incorrectly spelled letters.

Figure 2: Geographical Distribution of the Scientists and Engineers



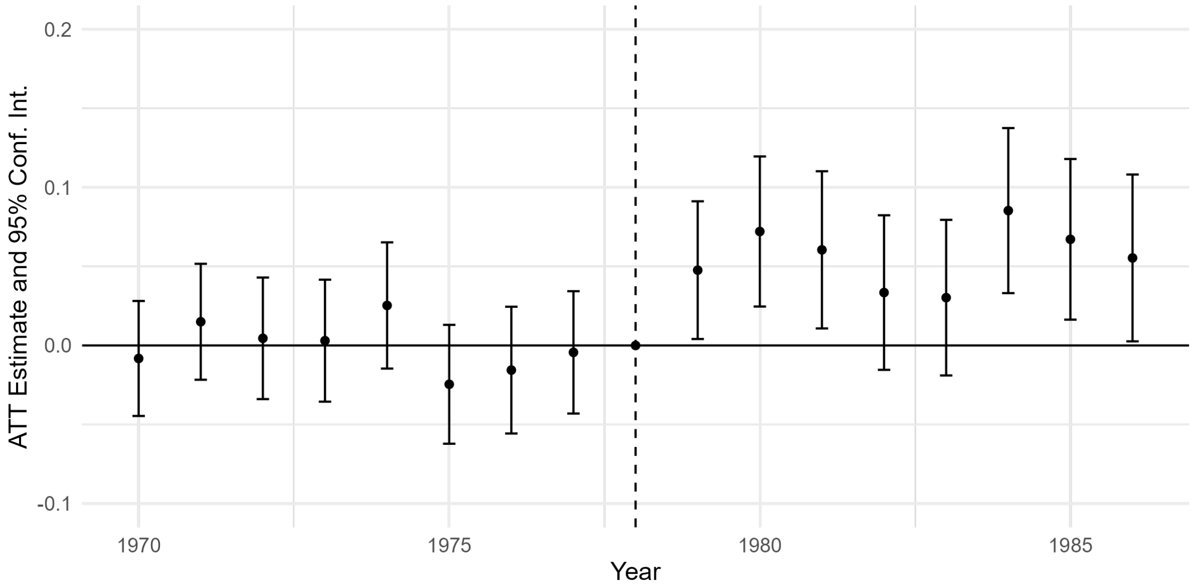
Notes: This figure plots the geographical distribution of the scientists, using county delineations from the 1990 census. The historical county FIPS crosswalk follows Eckert, Gvirtz, Liang, and Peters (2020). The scientist counts are weighted to account for differences in population weights between 1990 and 2010. For visualization purposes, the color scale is capped at 400. Counties with more than 400 scientists are represented using the same color as those with exactly 400 scientists.

Figure 3: Business Formation Trend in the U.S.



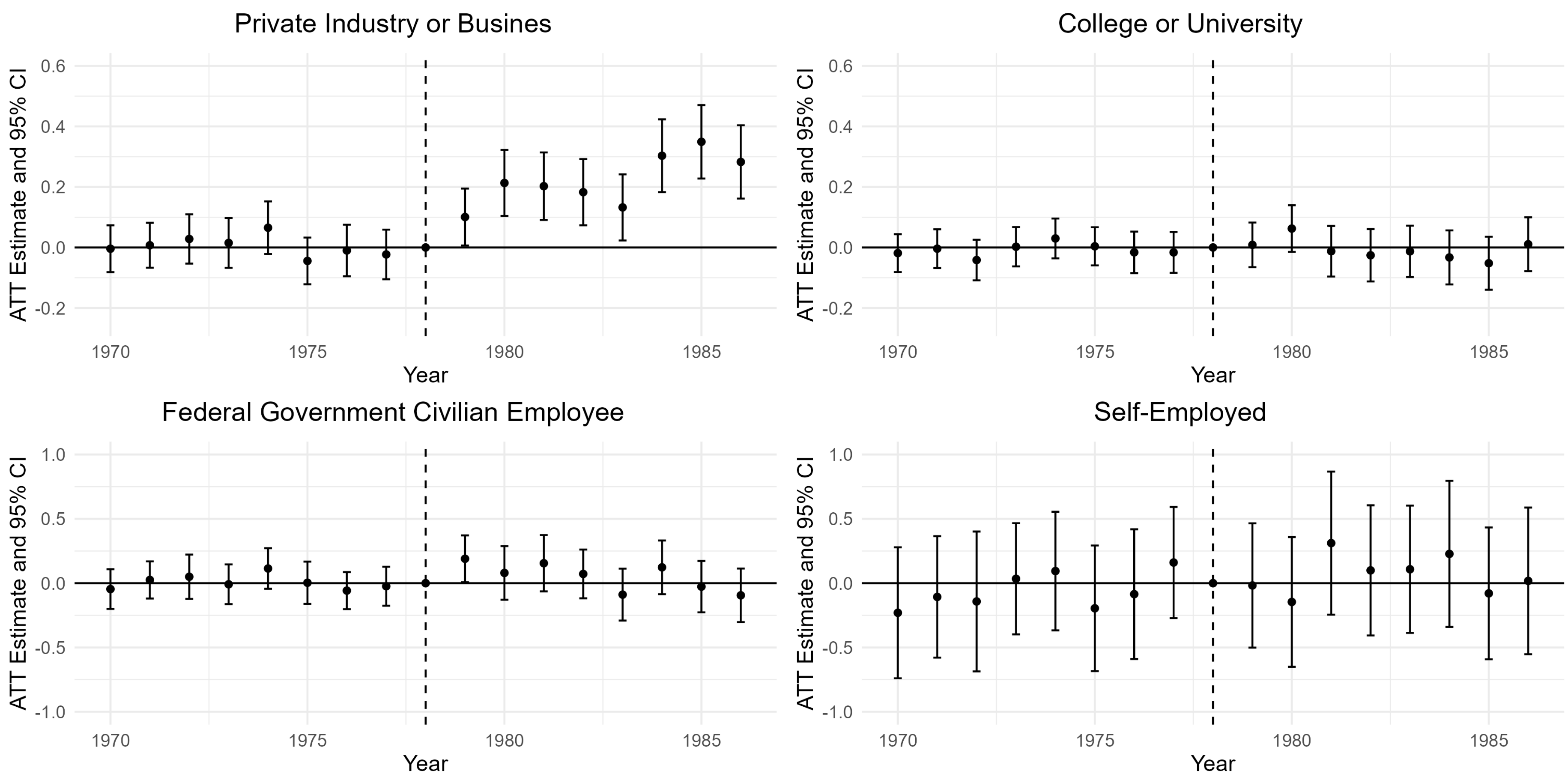
Notes: This figure plots the number of businesses incorporated in the U.S. and those founded by scientists. The data is from OpenCorporates. Business formation counts are normalized to 1978 (set to 1) for comparison. The total U.S. business formation includes all newly incorporated businesses, while scientist-founded businesses refer to firms established by scientists in my data sample. The data includes only business registrations where both the officers' names and company addresses are available.

Figure 4: Effects of VC on Scientists' Entrepreneurial Entry



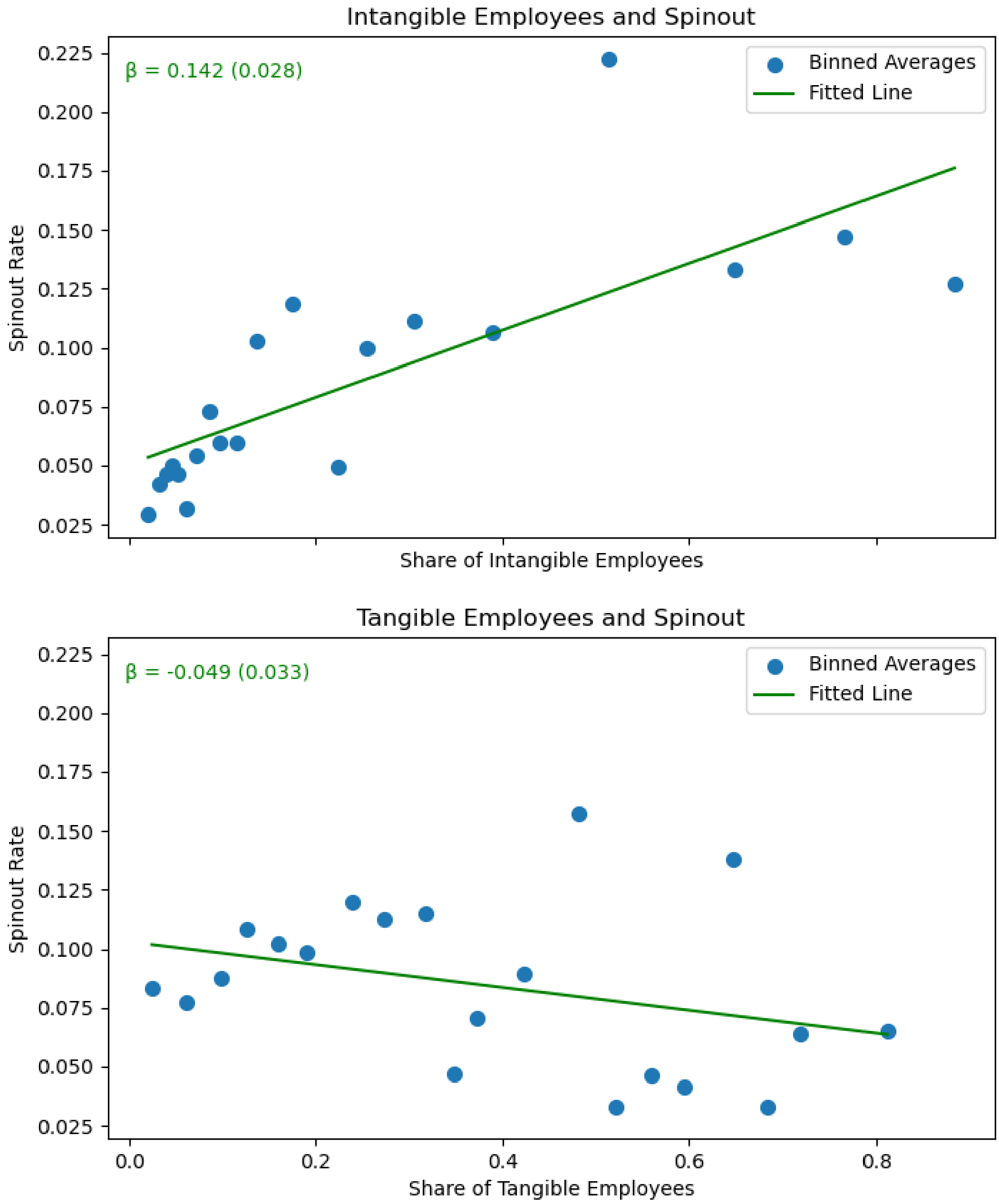
Notes: This figure displays the coefficient of year indicator variables interacted with the intangible specialty. The specification is from the difference-in-differences estimation from Column (4) in the Table 3. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 5: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Type of Employer



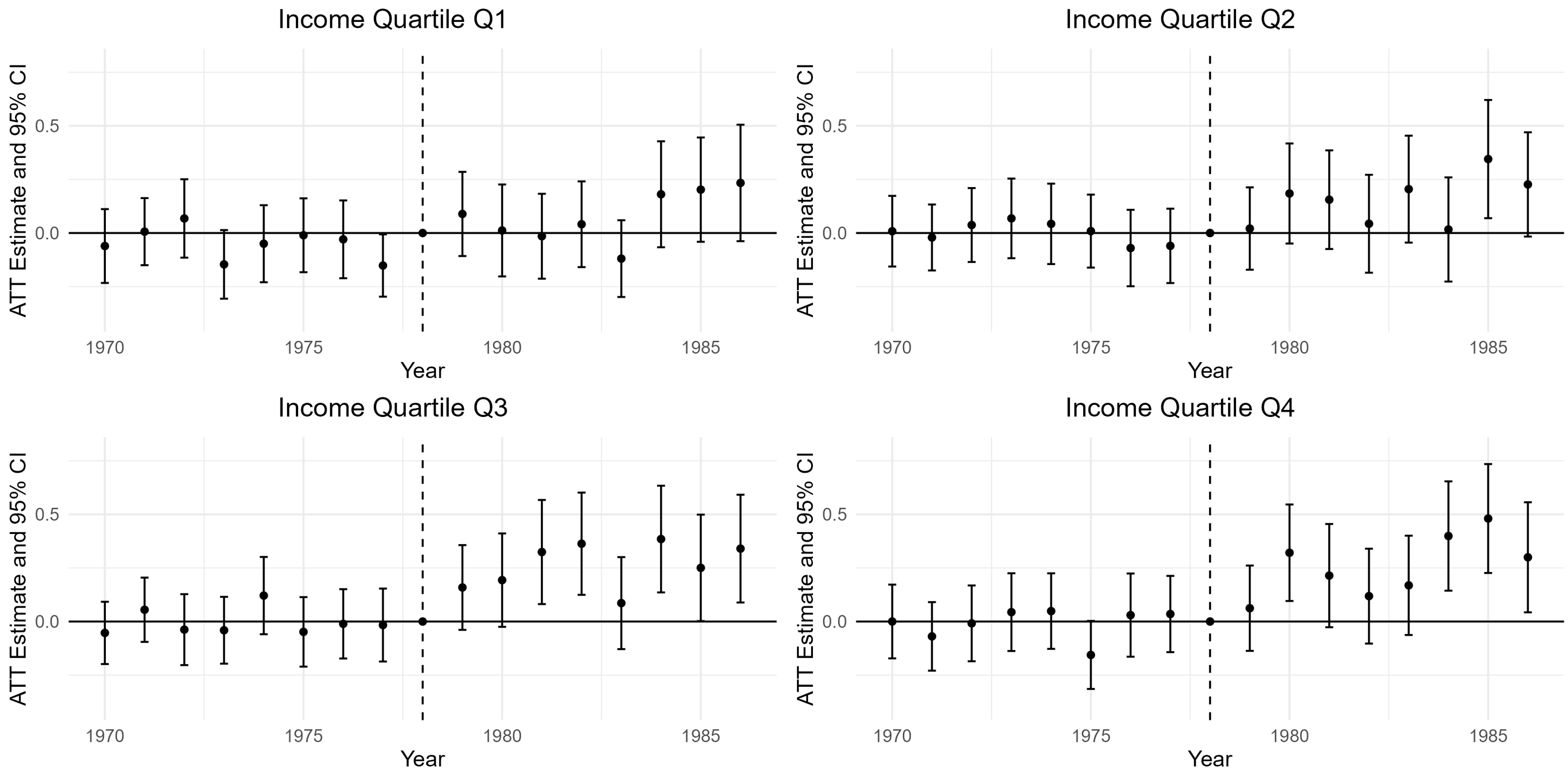
Notes: This figure displays the coefficients from the difference-in-differences estimation of Columns (1) - (4) in Table 4. It illustrates the heterogeneous treatment effects based on the type of employer for scientists and engineers. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 6: Intangible Employee Share and Spinout Rate at Firm Level



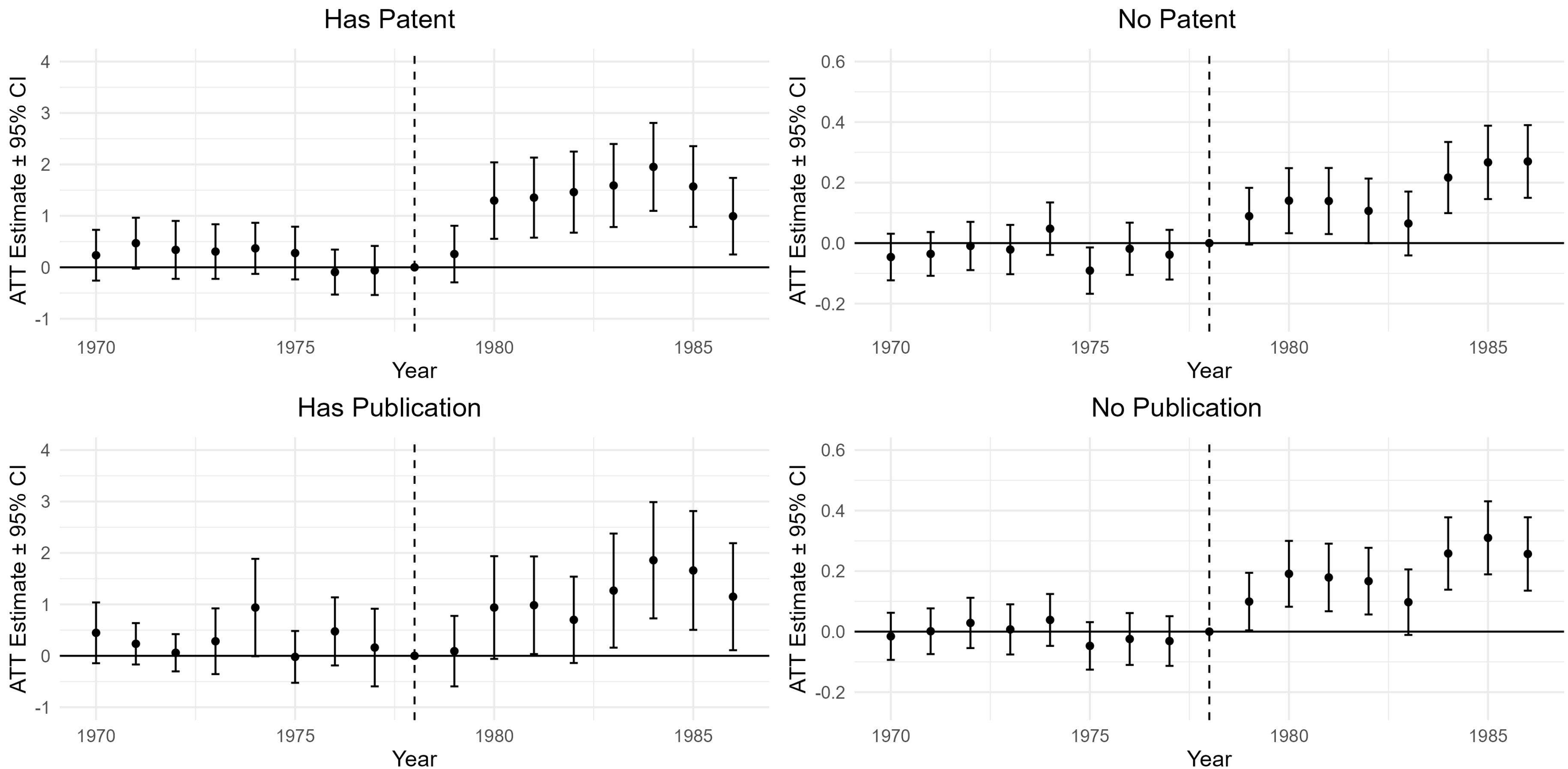
Notes: This figure displays the correlation between the spinout rate and its share of intangible (or tangible) employees at firm level. The share of intangible employees is defined as the number of intangible scientists divided by the total number of scientists in my data sample at the firm. Each dot shows the mean of firms whose observations fall within the corresponding bin. The green line represents the fitted values from a simple firm-level OLS regression with the spinout rate as the outcome variable and the share of intangible (or tangible) employees as the explanatory variable.

Figure 7: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Income



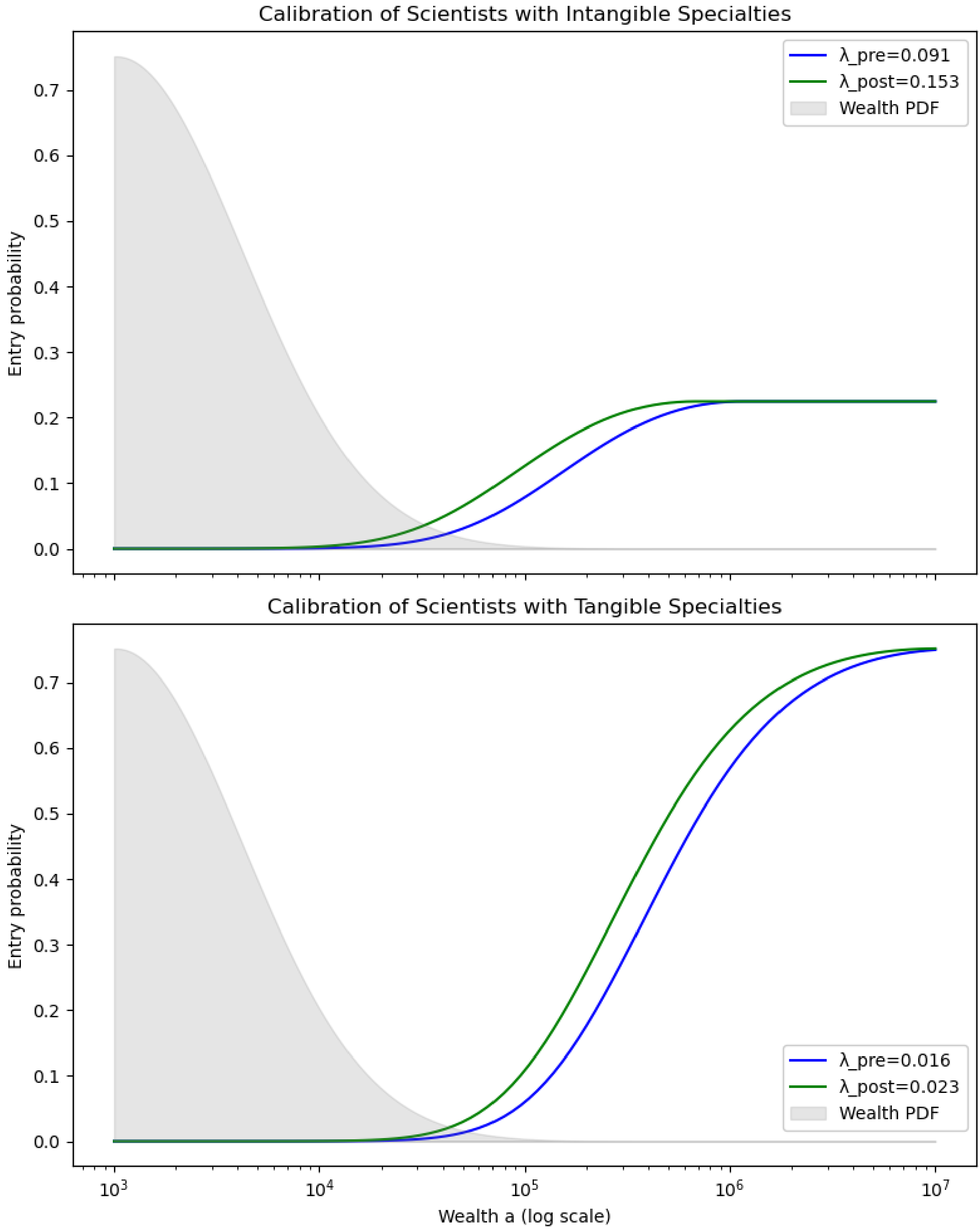
Notes: This figure displays the coefficients from the difference-in-differences estimations in Columns (1) - (4) in Table 5. It illustrates the heterogeneous treatment effects based on the type of employer of the scientists and engineers. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 8: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Employee Productivity



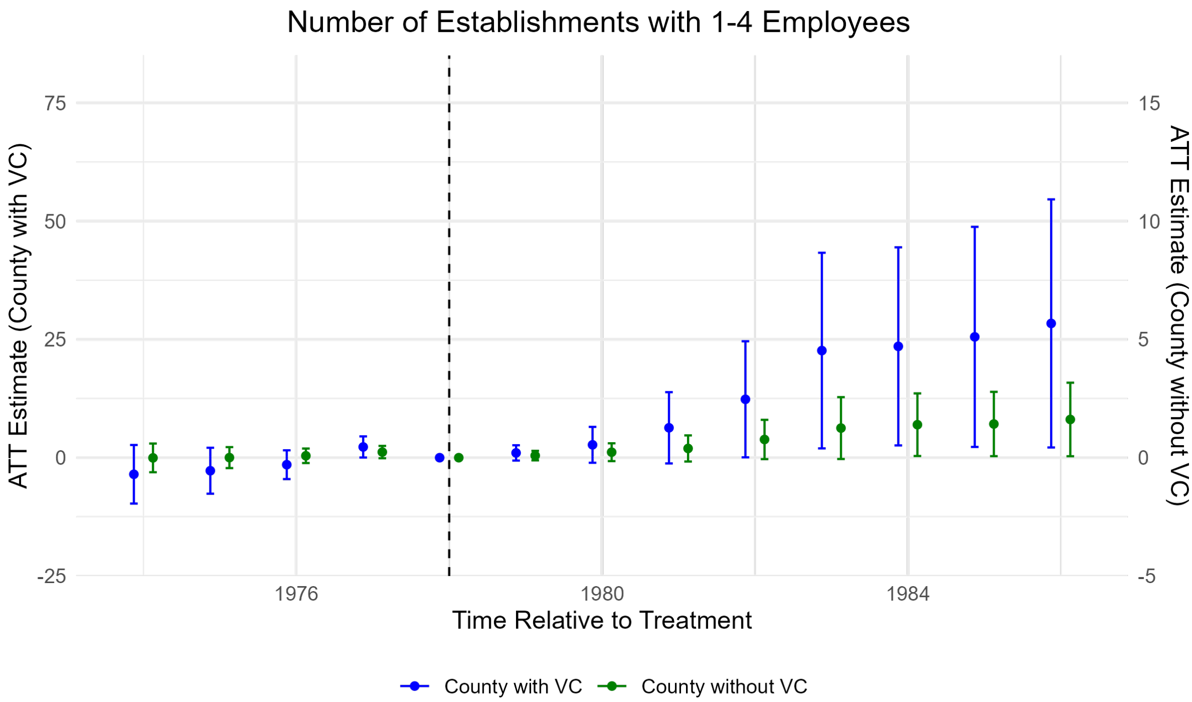
Notes: This figure displays the coefficients from the difference-in-differences estimations in Columns (1)-(4) in Table 6. It illustrates the heterogeneous treatment effects based on the patenting activity of the scientists and engineers. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 9: Financial Constraints and Business Formation Rate



Notes: This figure illustrates the implied entry probability as a function of wealth a, displayed on a logarithmic scale for the x-axis. The blue line represents λ_{pre} and the green line represents λ_{post}. ERISA relaxed the financial constraints and increased λ. The required parameters are calibrated to the values in Table 7.

Figure 10: Effects of VC on Industry Size



Notes: These figures display the coefficients from the difference-in-differences estimation in Table 8. The subfigure on the Number of Employees is from Columns (3) and (4). The subfigure on the Number of Establishments with 1-4 Employees is from Columns (7) and (8). The figures illustrate that the treatment effects of VC are precisely in the places where there was VC investment and which had limited spillover to other counties. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Tables

Table 1: Summary Statistics on Business Formation, Patenting, and Publication

Statistic	Count	Min	50%	Mean	95%	99%	Max	Std. Dev.
StartBusiness	458,703	0	0	0.03	0	1	1	0.17
BizCount	458,703	0	0	0.06	0	1	57	0.55
FilePatent	458,703	0	0	0.09	1	1	1	0.28
PatCount	458,703	0	0	0.37	1	8	356	2.75
HasPublication	458,703	0	0	0.10	1	1	1	0.29
PaperCount	458,703	0	0	2.12	7	53	1,273	15.11
Intangible Score	458,703	0.52	0.79	0.79	0.84	0.87	0.89	0.04
Tangible Score	458,703	0.51	0.78	0.78	0.89	0.95	0.95	0.05
Gross Income	370,120	100	12000	13186.12	25000	40000	99900	7719.46

Notes: This table presents the summary statistics of the variables related to the patenting and publication activities of scientists. All variables are at the individual level. BizCount represents the number of businesses formed by a scientist. StartBusiness equals one if a scientist started at least one firm. PatCount is the number of patents on which the scientist is listed as an inventor. FilePatent equals one if a scientist filed at least one patent. PaperCount is the number of journal publications authored by the scientist. HasPublication equals one if a scientist published at least one journal article. Intangible Score and Tangible Score are calculated based on the textual similarity between the work specialty of the scientists and the tangible and intangible dictionaries. Gross income is self-reported in the NRSTP.

Table 2: Top Intangible and Tangible Specialties

Tangible Specialties	Intangible Specialties
Mechanical engineering	Operations research
Electrical engineering	Theory and practice of computation
Chemical engineering	Mathematics of resource use
Plastics engineering	Operations analysis
Aerospace engineering	Communication science
Civil engineering	Project management and control
Textile engineering	Epidemiology
Materials engineering	Evolution
Electronics engineering	Information system design
Metallurgical engineering	Statistics

Notes: The table reports the top intangible and tangible scientific specialties based on textual similarity. Mechanical engineering has the highest difference between the tangible and intangible scores, indicating that it is highly tangible. In contrast, theory and practice of computation has the lowest difference between the tangible and intangible scores, suggesting it is the most intangible specialty.

Table 3: Effect of VC on Scientists' Entrepreneurial Entry

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
Post1979 × Intangible	0.0329***	0.0215***	0.0213***	0.0506***
	(0.0069)	(0.0069)	(0.0069)	(0.0074)
Intangible	0.0158***	0.0145***	0.0144***	
	(0.0038)	(0.0038)	(0.0038)	
Post1979	0.0602***	0.0395***		
	(0.0043)	(0.0042)		
Constant	0.0744***	0.0720***		
	(0.0025)	(0.0025)		
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	4,250,561	4,250,561	4,250,561	4,250,561
R^2	0.00015	0.00043	0.00045	0.15388

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 4: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Type of Employer

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
	Private Industry	College and University	Federal Government	Self Employed
Post1979 × Intangible	0.1828***	0.0053	0.0392	0.0376
	(0.0171)	(0.0125)	(0.0331)	(0.0775)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,563,067	1,262,682	200,679	75,748
R^2	0.14871	0.15071	0.14820	0.22156

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The sample is split based on the type of employer. The type of employer is divided into four categories: private industry when the scientist works for a for-profit firm, college and university, federal government civilian employee, and self-employed. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Productivity

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	Panel A: Scientists Working in Private Industry or Sector			
	(1)	(2)	(3)	(4)
	Q1	Q2	Q3	Q4
Post1979 × Intangible	0.0841***	0.0997***	0.2083***	0.2561***
	(0.0317)	(0.0332)	(0.0370)	(0.0368)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	373,070	331,704	340,646	341,314
R^2	0.15877	0.14569	0.13976	0.16011
	Panel B: Scientists Working for Colleges or Universities			
	(5)	(6)	(7)	(8)
	Q1	Q2	Q3	Q4
Post1979 × Intangible	-0.0357	-0.0009	0.0358	-0.0226
	(0.0240)	(0.0259)	(0.0269)	(0.0318)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	238,763	268,001	305,496	295,606
R^2	0.14527	0.13549	0.15381	0.15780

Notes: This table presents difference-in-differences estimates of the impact of the 1979 ERISA reform on business formation by scientists over the period 1970–1986. The sample is stratified by quartiles of self-reported gross income, defined separately for private-sector and university-employed scientists. For scientists in the private sector, the gross income quartile thresholds are \$100 (0th percentile), \$10,000 (25th), \$13,200 (50th), \$17,400 (75th), and \$99,900 (100th). For university scientists, the corresponding thresholds are \$100 (0th percentile), \$7,500 (25th), \$11,000 (50th), \$15,000 (75th), and \$99,900 (100th). The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6: Heterogeneous Effect on Scientists' Entrepreneurial Entry by Innovation Activity

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	Panel A: Scientists Working in Private Industry or Sector			
	(1)	(2)	(3)	(4)
	Has Patent	No Patent	Has Publication	No Publication
Post1979 × Intangible	0.7385***	0.1664***	0.3758***	0.1766***
	(0.1123)	(0.0171)	(0.1454)	(0.0171)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	289,261	1,273,806	30,589	1,532,478
R^2	0.14593	0.14937	0.14839	0.14851
	Panel B: Scientists Working for Colleges or Universities			
	(5)	(6)	(7)	(8)
	Has Patent	No Patent	Has Publication	No Publication
Post1979 × Intangible	0.8627***	0.0028	0.0092	0.0039
	(0.2252)	(0.0123)	(0.0325)	(0.0135)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	25,165	1,237,517	229,379	1,033,303
R^2	0.15228	0.15046	0.15573	0.14881

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The sample is split based on the type of employer and whether the scientists filed a patent or published a journal article before 1979. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7: Calibrated Parameters and Values

Parameter	Description	Data Source	Value
$\alpha_{\text{intangible}}$	Return to scale of the production process for intangible scientists	Calibration data of high-tech firms (primarily software and IT firms) from 1966 to 1984 (Crouzet and Eberly, 2023)	0.845
α_{tangible}	Return to scale of the production process for tangible scientists	Calibration data of manufacturing firms from 1966 to 1984 (Crouzet and Eberly, 2023)	0.940
r	Gross interest rate	10 Year Treasury Yield in January 1980	1.108
$w_{\text{intangible}}$	Average annual gross income for intangible scientists	NRSTP in 1965 and scaled to 1980 values	13767.31
w_{tangible}	Average annual gross income for tangible scientists	NRSTP in 1965 and scaled to 1980 values	13130.42
$\eta_{\text{intangible}}$	Share of intangible capital in production for intangible scientists	Calibration data of high-tech firms from 1966 to 1984 (Crouzet and Eberly, 2023)	0.324
η_{tangible}	Share of intangible capital in production for tangible scientists	Calibration data of manufacturing firms from 1966 to 1984 (Crouzet and Eberly, 2023)	0.176
μ_a	Mean of log wealth	Derived from net family assets in the 1976 survey (Evans and Jovanovic, 1989)	8.91
σ_a	Std. of log wealth	Derived from net family assets in the 1976 survey (Evans and Jovanovic, 1989)	1.41
μ_z	Mean of entrepreneurial ability	(Evans and Jovanovic, 1989)	2.00
σ_z	Std. of entrepreneurial ability	(Evans and Jovanovic, 1989)	0.90

This table presents the values used for the model calibration. Assume family assets X is log-normally distributed and $Y = \ln X$. Let μ and σ denote the mean and standard deviation of X. Then the variance of Y is $\sigma^2_{\ln} = \ln(1 + \sigma^2/\mu^2)$ and its mean is $\mu_{\ln} = \ln \mu - \frac{1}{2}\sigma^2_{\ln}$. With $\mu = 20{,}009.2$ and $\sigma = 50{,}053.3$ from the paper, I obtain $\sigma_{\ln} \approx 1.41$ and $\mu_{\ln} \approx 8.91$.

Table 8: Effect of VC on Industry Size

Panel A: Employment				
Dependent Variable:	**Number of Employees**			
	(1)	(2)	(3)	(4)
	Full Sample		VC County	Non VC County
Post1979 × Intangible	85.82***	106.2***	592.0***	28.70***
	(9.580)	(10.67)	(68.31)	(2.297)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes			
County FE	Yes			
Year-County FE		Yes	Yes	Yes
Observations	990,668	990,668	136,564	854,104
R^2	0.35723	0.36191	0.40690	0.22228

Panel B: Establishments				
Dependent Variable:	**All Sizes**		**1-4 Employees**	
	(5)	(6)	(7)	(8)
	VC County	Non VC County	VC County	Non VC County
Post1979 × Intangible	32.25***	1.847***	16.44***	0.8349***
	(3.445)	(0.0935)	(2.099)	(0.0558)
Industry FE	Yes	Yes	Yes	Yes
Year-County FE	Yes	Yes	Yes	Yes
Observations	136,564	854,104	136,564	854,104
R^2	0.40348	0.37004	0.35618	0.33564

Notes: This table reports the difference-in-differences estimates of the ERISA effect on industry size from 1974 to 1986. *VC County* represents a subsample with counties that had at least one VC investment during 1979-1986. *Non VC County* is the opposite. *Number of Employees* is the number of paid employees during the payroll period. *Establishments* is the number of establishments with paid employees, categorized by employment-size class (e.g., 1–4 employees). *Intangible* is a binary variable indicating whether the industry is classified as intangible. *Post1979* equals one for years after 1978. Standard errors are clustered at the county level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix for

"Venture Capital and Scientists' Selection into Entrepreneurship"

Xuelai Li

September 2025

A Construction of Specialty Dictionaries

The tangible and intangible specialty dictionaries are constructed using GPT (including o3 and o4-mini). Firms are double-sorted using COMPUSTAT data based on the share of intangible assets and the capital expenditure-to-assets ratio. Company descriptions and financials of the top 100 firms at each extreme are extracted and saved as separate files. The GPT prompt is:

```
You are provided with three files:
tangible_companies.csv - Descriptions of the top tangible companies of
the 1980s.
intangible_companies.csv - Descriptions of the top intangible companies
of the 1980s.
specialty.csv - A list of scientists' specialties, one specialty per
line.


Task
1. From specialty.csv, select 20 specialties most relevant to
tangible_companies.csv.
2. From specialty.csv, select 20 specialties most relevant to
intangible_companies.csv.
3. Base the relevance on how closely each specialty aligns with the
companies' technologies and products.
4. Ensure all 40 chosen specialties are unique (no duplicates across the
```

```
two lists).


Output
Return only the following Python lists, without comments or explanations.
```

B Model

First consider the choice of k_1 and k_2 for the entrepreneur. F.O.C. gives:

$$\frac{k_2}{k_1} = \frac{\eta}{1-\eta} \quad \Longrightarrow \quad k_1 = (1-\eta)I, \;\; k_2 = \eta I.$$

Hence the pledgeability constraint implies

$$I \;\leq\; \lambda a + b \;\leq\; \lambda a + \phi(1-\eta)I \quad \Longrightarrow \quad I \;\leq\; I_{\max} \;\equiv\; \frac{\lambda a}{1-\phi(1-\eta)}.$$

Define $\kappa \equiv \;\left[\eta^{\eta}(1-\eta)^{1-\eta}\right]^{\alpha}$ for simplicity. The problem reduces from finding k_1 and k_2 to only finding I:

$$\max_{0 \leq I \leq I_{\max}} \; \pi(I) \;=\; z\,\kappa I^{\alpha} - r\,(I-a).$$

$$F.O.C \qquad \alpha\kappa z I^{\alpha-1} - r = 0$$

The optimal I^* is $(\frac{\alpha\kappa z}{r})^{\frac{1}{1-\alpha}}$

The entrepreneur will be unconstrained whenever the optimal I^* is below $I_{\max}$, i.e.

$$z \leq \frac{r}{\alpha\kappa}(\frac{\lambda a}{1-\phi(1-\eta)})^{1-\alpha}$$

Solving for the ability threshold below and above $\frac{\lambda a}{1-\phi(1-\eta)}$ respectively:

$$w^{1-\alpha}\left(\frac{r}{\alpha\kappa}\right)^{\alpha}(1-\alpha)^{\alpha-1} \leq z \leq (\frac{\lambda a}{1-\phi(1-\eta)})^{1-\alpha}\left(\frac{r}{\alpha\kappa}\right) \tag{1'}$$

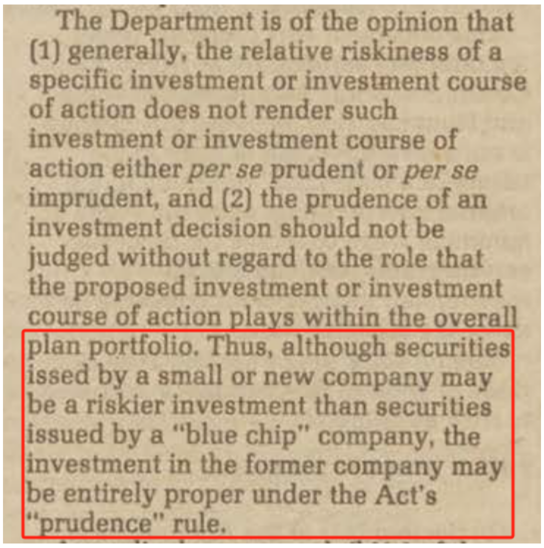
or

$$z > \max\left[\left(\frac{\lambda a}{1-\phi(1-\eta)}\right)^{1-\alpha} \left(\frac{r}{\alpha\kappa}\right),\ w\left(\frac{\lambda a}{1-\phi(1-\eta)}\right)^{-\alpha} + r\left(\frac{\lambda a}{1-\phi(1-\eta)}\right)^{1-\alpha} \right] \quad (2')$$

If z satisfies either constraint, the individual chooses entrepreneurship.

C Appendix Figures

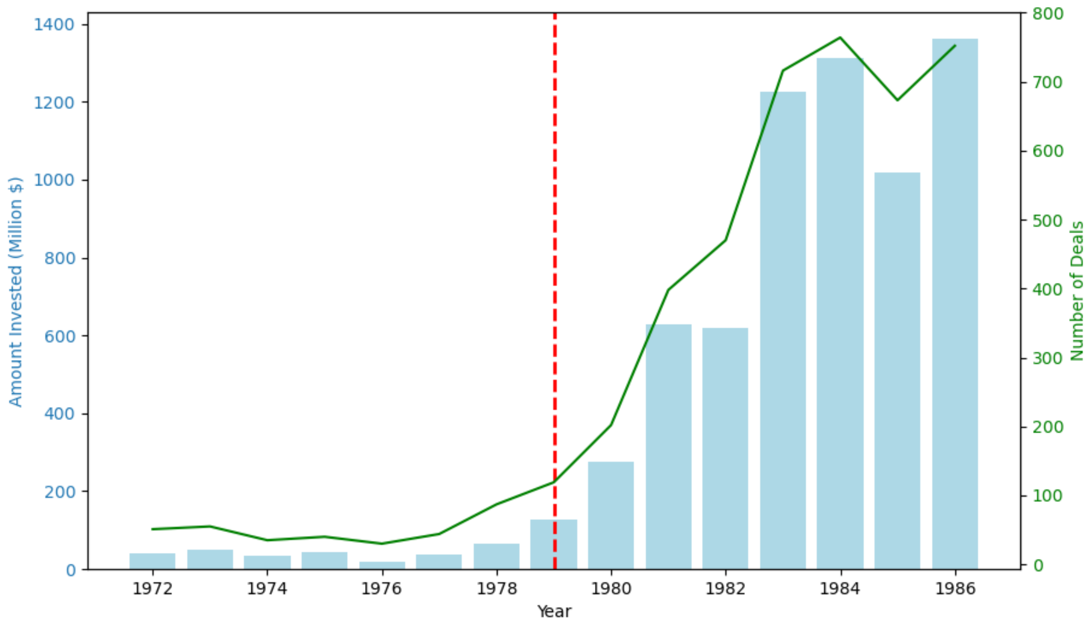
Figure A1: Original Document of the ERISA Reform in 1979



The Department is of the opinion that (1) generally, the relative riskiness of a specific investment or investment course of action does not render such investment or investment course of action either *per se* prudent or *per se* imprudent, and (2) the prudence of an investment decision should not be judged without regard to the role that the proposed investment or investment course of action plays within the overall plan portfolio. Thus, although securities issed by a small or new company may be a riskier investment than securities issued by a "blue chip" company, the investment in the former company may be entirely proper under the Act's "prudence" rule.

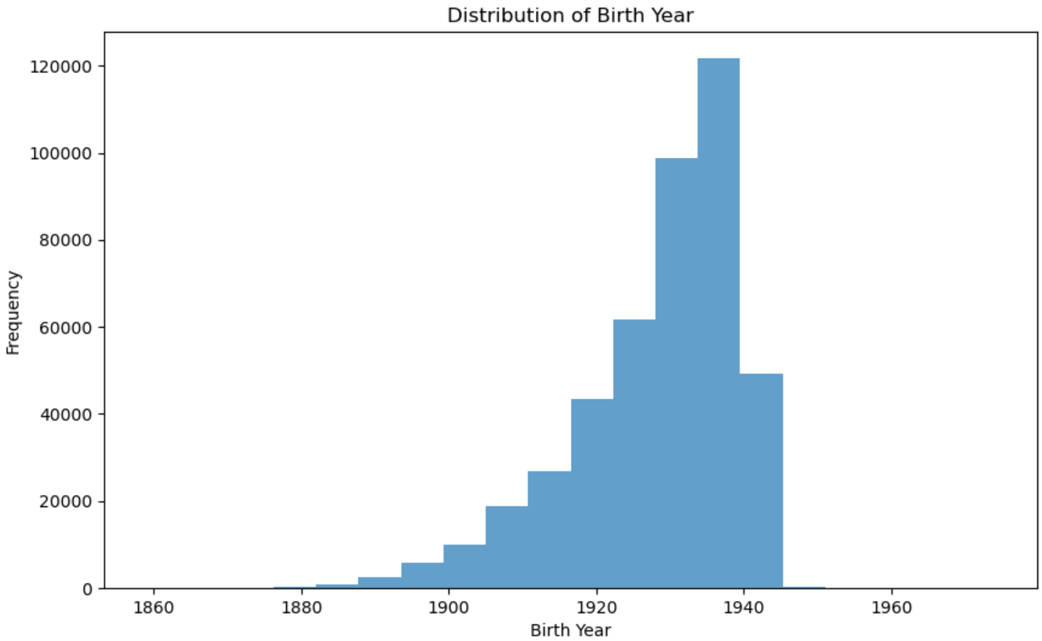
Notes: This graph shows Title 29 of the U.S. Code of Federal Regulations Part 2550 of 1979. This is the final regulation on the "Rules and Regulations for Fiduciary Responsibility; Investment of Plan Assets Under the 'Prudence' Rule". The amendment was published in the Federal Register on June 26, 1979. Federal agencies typically begin drafting amendments well before the public discussion. The discussions within the Department of Labor (DOL) regarding fiduciary investment duties likely started as early as 1978. The DoL would publish a Notice of Proposed Rulemaking (NPRM) in the Federal Register to inform the public of the proposed changes and invite comments. This step often occurs 6–18 months before the final rule is published. For the § 2550.404a-1 amendment, the NPRM likely appeared in the Federal Register in late 1978 or early 1979. Following the NPRM, there would have been a public comment period (typically 30–90 days) during which stakeholders could provide feedback. After the comment period, the DOL would review the feedback, potentially revise the proposal, and prepare the final rule for publication.

Figure A2: VC Investment and ERISA Reform



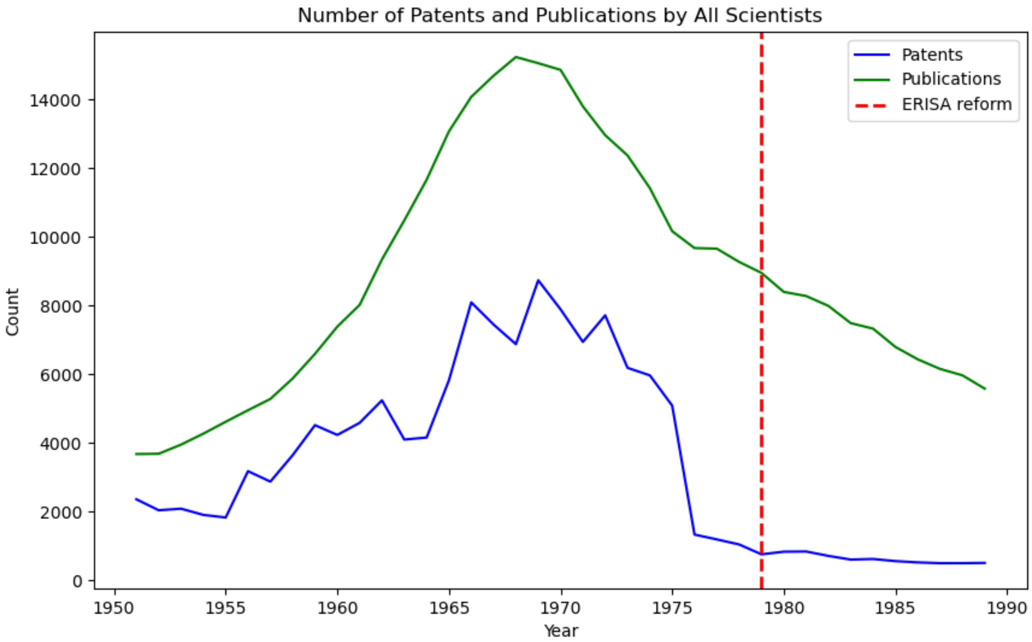
Notes: This figure plots the total amount of VC investment and the number of deals in the U.S. Note that these values are underestimated due to incomplete data coverage in the dataset, as many of the deals did not disclose the deal sizes. The data comes from Venture Economics, a database focusing on the venture capital and private equity sectors. The database includes fields such as investors, invested startups, and fund profiles. This is the only database that covers the VC and PE deals in 1970s, making it a valuable resource for the analysis in this study. Many foundational papers in the entrepreneurial finance literature use this database (Kortum and Lerner, 2000; Ewens, Nanda, and Rhodes-Kropf, 2018).

Figure A3: Birth Year of the Scientists



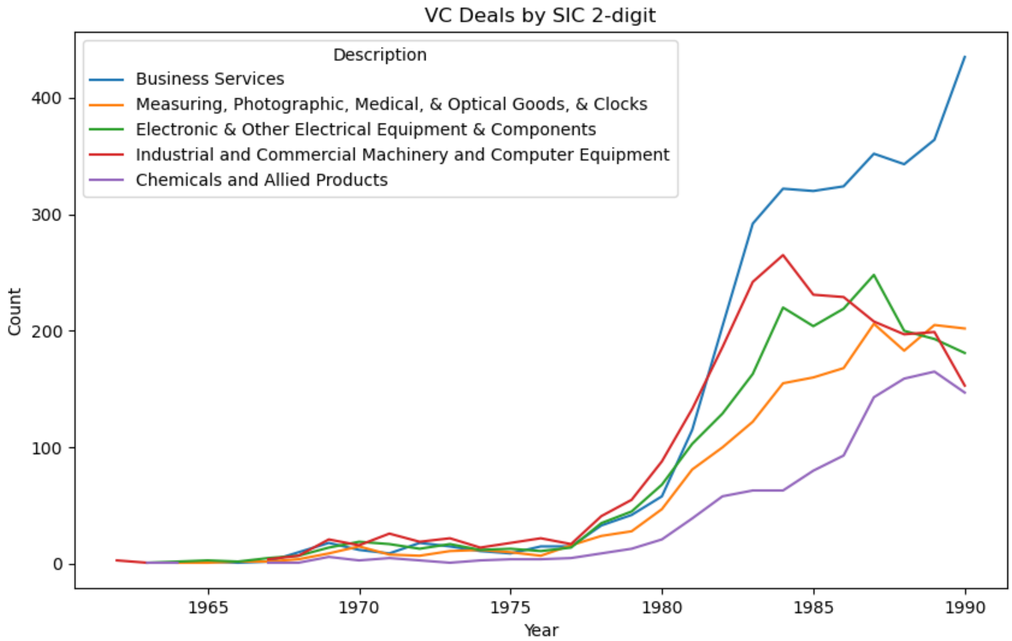
Notes: This figure plots the distribution of scientists' birth years. The birth year is self-reported in the AMS data. Since this information is not reported in the NRSTP data, the birth year of scientists recorded in the NRSTP is calculated based on the year and level of the highest degree.

Figure A4: Patents and Publications by Scientists



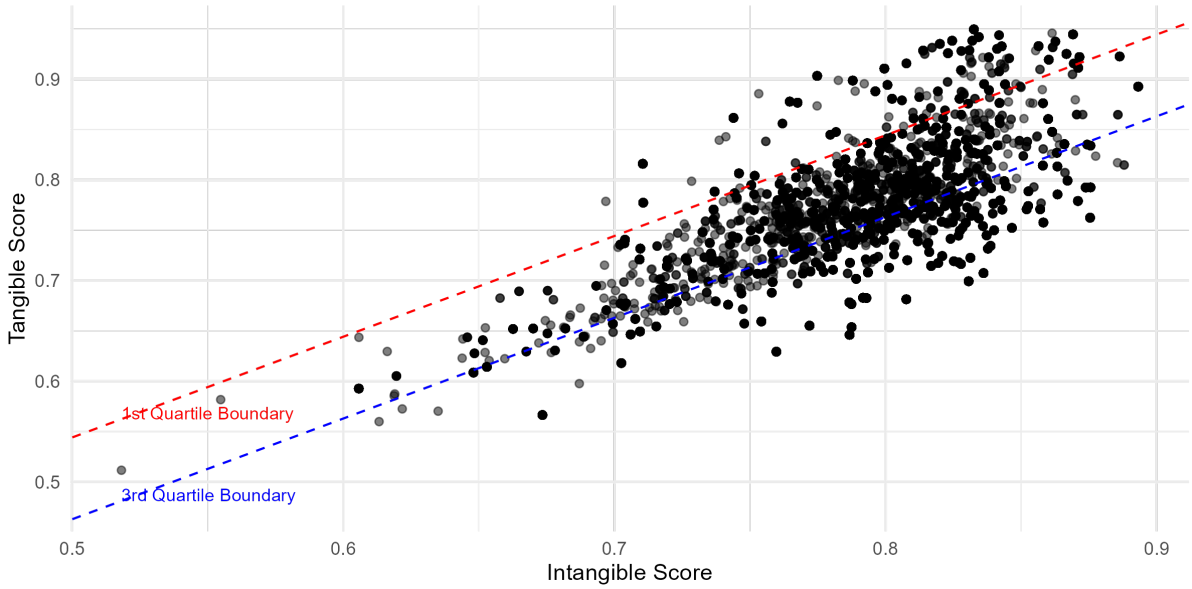
Notes: This figure plots the number of granted patents filed and published papers by scientists in my data sample over the years. The dataset includes only patents that were granted; applications that did not result in a grant are not observed. Only papers published in journals are included as publications.

Figure A5: Early-Stage VC Deals by Industry



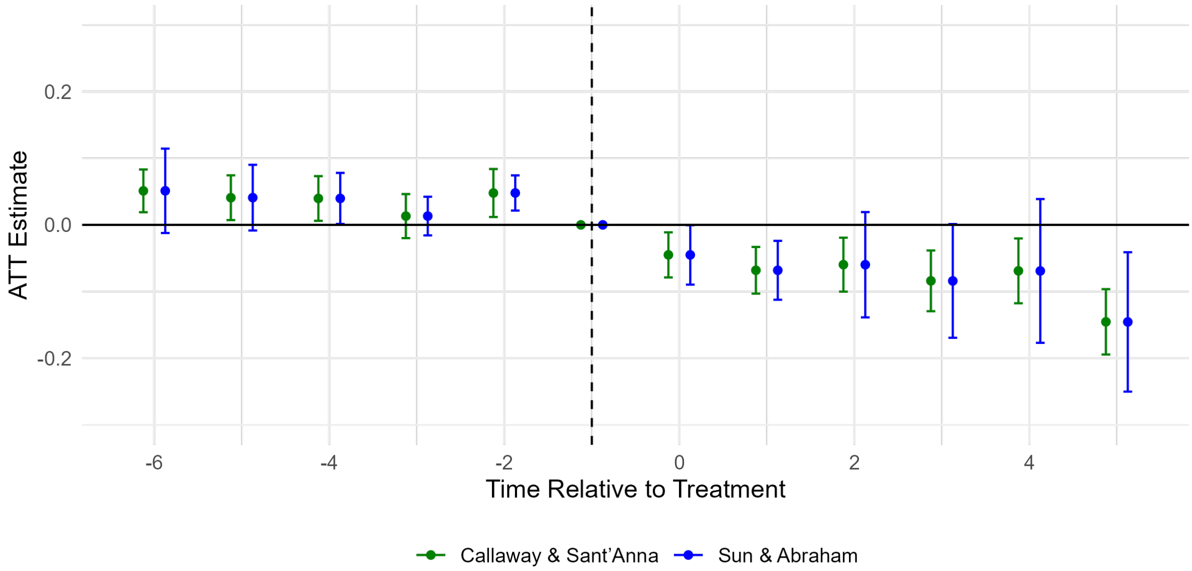
Notes: This figure plots the number of VC deals from 1960 to 1990 based on the two-digit SIC codes. The top five industries by deal count in 1990 are selected. The data is from Venture Economics.

Figure A6: Difference between Tangible and Intangible Scores



Notes: This figure plots the tangible and intangible scores for each specialty. Each point represents a unique scientist, with their scores derived from their work specialty. The 45-degree dashed lines represent boundaries where the difference between the intangible and tangible scores equals the first and third quartiles of the empirical distribution. The red line corresponds to the 75th percentile of the difference (intangible minus tangible), while the blue line denotes the 25th percentile.

Figure A7: Negative Credit Shock to Scientists' Entrepreneurial Entry



Notes: This figure plots the estimated coefficients from the event study based on Equation $BusinessFormation_{ist} = \beta Deregulation_{st} + \eta_i + \eta_t + \epsilon_{ist}$. $Deregulation_{st}$ is a dummy variable that equals one in the year following the implementation of intrastate banking deregulation in a given state. Since intrastate banking deregulation included both M&A and de novo deregulation, I follow the previous literature (Chava et al., 2013; Jayaratne and Strahan, 1996) in classifying a state as "intrastate deregulated" in the year after either M&A or de novo deregulation occurred. The green lines represent estimates based on Callaway and Sant'Anna (2021), which report group-time average treatment effects using never-treated units as the control. The blue lines are based on Sun and Abraham (2021), which present event-time treatment effects using never-treated or not-yet-treated units as the control. The results highlight the heterogeneous effects of intrastate bank deregulation on business formation across different specialties. The vertical lines denote 95% confidence intervals for the coefficient estimates. The results indicate a negative impact of intrastate bank deregulation on business formation among scientists with tangible specialties. Specifically, the estimated average treatment effect (ATT) of scientists is -0.137, with a standard error of 0.0167 based on Callaway and Sant'Anna (2021).

D Appendix Tables

Table A1: Level and Year of Highest Degree by Data Source

Level of Highest Degree	Count
Bachelor	142,242
Master's	130,488
MD	10,996
PhD	160,082
PhD+	2,336

Notes: Compared to AMS, the NRSTP offers a broader view of the workforce. The NRSTP covers a wider range of fields and is more oriented toward workforce analysis, while AMS emphasizes individual recognition and contributions within the scientific community. AMS primarily includes renowned scientists, most of whom are affiliated with universities and hold PhDs. In contrast, the NRSTP encompasses a broader group of individuals engaged in R&D activities, many of whom may not possess advanced degrees. PhD+ means that the person has more than one PhD degree, or has both PhD and MD degrees.

Table A2: Institution of Highest Degree

University of Highest Degree	Count
University of Michigan-Ann Arbor	10,480
Columbia University in the City of New York	10,060
Harvard University	9,728
University of California-Berkeley	8,302
New York University	7,756
Purdue University	7,529
University of Wisconsin	7,499
Ohio State University	7,378
Massachusetts Institute of Technology	7,287
University of Chicago	7,142

Notes: This table reports the institution of the highest degree of the scientists and engineers in my sample. Universities within the University of California system have missing values because many records only include the UC system but do not specify the specific campus attended.

Table A3: Average Pre-Tax Income by Income Quantiles ($ 2018)

Quantile	NRSTP	PSZ
Bottom 50%	42,061	13,761
Middle 40%	81,616	40,050
Top 10%	124,817	132,719
Top 5%	164,118	193,714
Top 1%	249,437	472,005
Top 0.5%	344,038	687,512
Top 0.001%	520,178	20,274,790

Notes: This table shows the scientists' income distribution and compares it with that of the general U.S. population. The PSZ data is from the 2022 version of TB3 from Distributional National Accounts by Piketty, Saez, and Zucman (2018): https://gabriel-zucman.eu/usdina/. Last retrieved on September 23, 2025.

Table A4: Type of Employer

Employment Sector	Count
Private Industry or Business	171,484
College or University	138,280
State, Local, or Other Government (except educational institution)	39,647
Federal Government Civilian Employee	27,968
Other Educational Institution	15,421
Military Service, Active Duty	11,529
Nonprofit Organization	10,913
Self-Employed	9,162
Other	2,167

Notes: This table reports the types of employers for the scientists and engineers in my sample. Over the years, the classification of employer type has become increasingly granular. I manually create a crosswalk file to harmonize these classifications. In 1970, the category "State, local, or other government (except educational institution)" included entities such as the USPHS Commissioned Corps, U.S. Weather Bureau, State Government, International Agencies, and Other Government Agencies. Research centers managed by for-profit organizations are classified under "Private Industry or Business," while those managed by educational institutions are classified as "College or University."

Table A5: Top Employers of Scientists and Engineers

Firm Name	NAICS Industry Name	Count
DuPont de Nemours, Inc.	Chemical Manufacturing	4,792
International Business Machines	Computer and Electronic Product	3,198
Union Carbide Corp	Chemical Manufacturing	3,110
General Electric Company	Electrical Equipment	2,674
Shell Oil Co.	Petroleum and Coal Products	2,228
Dow Chemical Company	Chemical Manufacturing	2,011
Monsanto Co	Chemical Manufacturing	1,653
Humble Oil & Refining Co	Petroleum and Coal Products	1,348
North American Rockwell	Aerospace Product and Parts	1,311
Eastman Kodak Co	Photographic and Optical Equipment	1,165
Mobil Oil Corp	Petroleum and Coal Products	1,130
Lockheed	Aerospace Product and Parts	1,095
Texaco Inc	Petroleum and Coal Products	1,093
Allied Chemical Corp	Chemical Manufacturing	1,089
Esso Chem Co Inc	Chemical Manufacturing	1,065
Westinghouse Electric Corp	Electrical Equipment and Component	1,035
Phillips Petroleum Co.	Petroleum and Coal Products	990
American Cyanamid Co	Chemical Manufacturing	976
Bell Telephone Company	Telecommunications	971
Boeing Company	Aerospace Product and Parts	948
Radio Corporation of America	Broadcasting and Communications	928
Gulf Oil Corp	Petroleum and Coal Products	857
Chevron Corporation	Petroleum and Coal Products	847
Hercules Inc	Chemical Manufacturing	840
3M Company	Miscellaneous Manufacturing	705
Battelle Memorial Institute	Research and Development Services	692
McDonnell Douglas Aircraft	Aerospace Product and Parts	688
Standard Oil Co	Petroleum and Coal Products	688
Pan American World Airways	Air Transportation	673
Sperry Rand Corp	Computer and Electronic Product	671

Notes: This table shows the top employers of the scientists and engineers in my sample. I standardize and consolidate information on mergers and acquisitions (M&As) by aligning historical corporate entities with their post-merger counterparts. Firms that merged before 1972, such as North American Rockwell Corporation (1967) and McDonnell Douglas Aircraft Corporation (1967), were identified and recorded to maintain historical accuracy. Similarly, post-1972 M&As, including Lockheed Martin Corporation (1995) and Northrop Grumman Corporation (1994), were documented by tracing their predecessor firms.

Table A6: First Specialty of Work

Specialty	Count
Organic Chemistry	47,178
Agricultural and Biological Sciences	36,514
Geology	23,054
Analytical Chemistry	18,728
Physical Chemistry	16,581
Related Chemical Specialties	13,842
Theory and Practice of Computation	13,733
Clinical Psychology	11,740
Biochemistry	11,288
Inorganic Chemistry	8,661
Chemistry	7,751
Probability and Statistics	6,857
Chemical Engineering	6,828
Solid State Physics	6,556
Nuclear Physics	5,350
Forestry	4,862
Optics	4,836
Civil Engineering	4,822
Mathematics of Resource Use	4,801
Electronics	4,580

Notes: This table reports the work specialties of the scientists and engineers in my sample. The data comes from both NRSTP and AMS. The NRSTP data originates from the "Professional Characteristics" section of the questionnaire, where respondents were asked to identify the specialties in which they believed they had demonstrated professional competence in research. While the classification of work specialties aligns with the categorization of academic majors, it provides a more detailed structure, incorporating multiple hierarchical levels of specialties for greater granularity. The AMS data comes from the list of academic disciplines provided by AMS.

Table A7: Correlation Matrix

	StartBusiness	FilePatent	HasPublication	Tangible Score	Intangible Score	Gross Income	isMale
StartBusiness	1.000						
FilePatent	0.048	1.000					
HasPublication	0.019	-0.017	1.000				
Tangible Score	0.004	0.075	-0.076	1.000			
Intangible Score	0.018	-0.123	-0.027	0.361	1.000		
Gross Income	0.058	0.165	0.099	-0.037	-0.012	1.000	
isMale	0.032	0.079	0.006	0.114	0.005	0.143	1.000

Notes: This table presents the correlation matrix between key variables in Table 1 and gender information. All variables are at the individual level. StartBusiness equals one if a scientist started at least one firm. FilePatent equals one if a scientist filed at least one patent. HasPublication equals one if a scientist published at least one journal article. Intangible Score and Tangible Score are calculated based on the textual similarity between the work specialty of the scientists and the tangible and intangible dictionaries. Gross income is self-reported in the NRSTP. Gender is either self-reported or guessed based on the first name.

Table A8: Dictionary of Tangible and Intangible Specialties with GPT o3

Tangible Specialties	Intangible Specialties
Mechanical Engineering	Business Finance and Administration
Electrical Engineering	Industrial Organization
Industrial Engineering	Economic Systems
Chemical Engineering	Theory and Practice of Computation
Materials Engineering	Business Data Processing
Metallurgical Engineering	Computer Science
Civil Engineering	Information Science
Aerospace Engineering	Patent Law
Mining and Petroleum Engineering	International Law
Food Science and Technology	International Economics
Forest Products	Project Management and Control
Food Packaging	Industrial and Personnel Psychology
Agricultural Engineering	Social Change and Development
Electronic Engineering	Economic Growth and Development
Environmental Engineering	General Economics
Process Engineering	Operations Research
Product Engineering	Systems Engineering
Industrial Hygiene	Demography and Population
Sanitary Engineering	Computer Hardware Design
Communications Engineering	Communications

Notes: The table reports the specialty dictionaries constructed using GPT o3 based on company descriptions from COMPUSTAT. Firms are double-sorted by their share of intangible assets and the capital expenditure-to-assets ratio. The top 100 firms at each extreme are used to generate the tangible and intangible dictionaries. Some specialty names are abbreviated for formatting purposes.

Table A9: Dictionary of Tangible and Intangible Specialties with GPT o4-mini

Tangible Specialties	Intangible Specialties
Mechanical Engineering	Communications Engineering
Electrical Engineering	Research Administration
Chemical Engineering	Information Science
Civil Engineering	Information System Design
Industrial Engineering	Information Retrieval
Materials Engineering	Computer Science
Metallurgical Engineering	Computer Hardware Design
Polymer Science	Business Organization
Ceramics	Management
Geology	Project Management and Control
Geophysics	Business Data Processing
Petroleum Engineering	Theory and Practice of Computation
Food Science and Technology	Probability and Statistics
Biomedical Engineering	Land Economics
Electronics	International Economics
Solid State Physics	Labor Economics
Electricity and Magnetism	Economic Growth and Development
Engineering Mechanics	Welfare Programs
Design Engineering	Hospital Administration
Environmental Engineering	Industrial Organization

Notes: The table reports the specialty dictionaries constructed using GPT o4-mini based on company descriptions from COMPUSTAT. Firms are double-sorted by their share of intangible assets and the capital expenditure-to-assets ratio. The top 100 firms at each extreme are used to generate the tangible and intangible dictionaries. Some specialty names are abbreviated for formatting purposes.

Table A10: Scientists' Work Specialty Tangibility Status (1962 vs. 1968)

	Intangible in 1968	
Intangible in 1962	0	1
0	7,709	138
1	1,050	6,034

Notes: This table presents the confusion matrix comparing the work specialties of the same scientists who appear in both the 1962 NRSTP survey and the 1968 survey. Scientists may change their work specialty over time, but the tangibility of each work specialty is time-invariant. The values represent the counts of observations transitioning between categories. Scientists whose work specialty changed from 1 to "not able to define" were dropped.

Table A11: Differences between Tangible and Intangible Scientists

Variable	Tangible	Intangible	Diff in Mean	t-statistic
Female	0.146	0.064	0.082	66.703
Year of Highest Degree	1955.837	1952.907	2.930	67.858
Basic Salary	13037.608	12925.898	111.710	4.042
Gross Income	13959.789	13264.311	695.478	20.432
Govt. Agriculture	0.022	0.068	-0.047	-58.263
Govt. Atomic Energy	0.019	0.028	-0.009	-14.724
Govt. Defense	0.116	0.072	0.044	37.414
Govt. Education	0.096	0.040	0.057	55.596
Govt. Natural Resources	0.011	0.038	-0.026	-43.588
Govt. Space	0.047	0.033	0.014	17.750
EmployerFirm	0.232	0.482	-0.250	-136.527
EmployerGov	0.057	0.039	0.018	20.633
EmployerMil	0.014	0.015	-0.001	-1.526
EmployerUni	0.417	0.196	0.221	122.622

Notes: The table reports the average differences between scientists with tangible and intangible specialties. *Basic Salary* and *Gross Income* are self-reported in the NRSTP. *Govt. Agriculture* indicates sponsorship by government agriculture programs, with similar definitions for Govt. Atomic Energy, Govt. Defense, Govt. Education, Govt. Natural Resources, and Govt. Space. *EmployerFirm* refers to scientists employed by private industry or business. *EmployerGov* denotes federal government civilian employees. *EmployerMil* represents military service personnel, and *EmployerUni* includes those in active duty at colleges or universities.

Table A12: Employers with the Top Shares of Scientists with Intangible and Tangible Work Specialties

Intangible Specialties	Tangible Specialties
Informatics Inc	Climax Molybdenum Co
Wyatt Co	Fritzsche Brothers Inc
Applied Data Research Inc	Detrex Chemical Industries Inc
Brookings Inst	Dexter Corp
Milliman & Robertson Inc	Sonoco Products Co
Computer Assoc Inc	Richardson Co
Scientific Data Systems	Schenectady Chem Inc
Computer Control Co	Drew Chemical Corp
Philip Hankins & Co Inc	Homestake Mining Co
Computing & Software Inc	Cosden Oil & Chem Co
Arthur Andersen & Co	Pennzoil Co
American Inst for Research	Ashland Oil
Touche Ross Bailey & Smart	Congoleum Nairn Inc
Data Dynamics, Inc.	Devoe & Raynolds Co Inc
Pacific Mutual Life Insurance Co	Travenol Labs Inc
Humrro	Westreco Inc
Austen Riggs Center	Fiberite Corp
Computer Usage Co	Neville Chem Co
Keystone Computer Assoc Inc	Holston Defense Corp
California Computer Products	H Kohnstamm & Co Inc

Notes: The table reports employers with the highest share of scientists specializing in either tangible or intangible fields. The share is calculated as the proportion of scientists with a tangible specialty relative to the total number of scientists. Employers are identified based on the workplace reported by scientists when completing the NRSTP or AMS survey.

Table A13: Top Frequent Words in the Business Names of Startups by Intangible and Tangible Scientists

Intangible Specialties	Tangible Specialties
music	oil
support	laboratories
data	petroleum
design	electric
foods	gas
rentals	furniture
steel	temple
knolls	scientific
communication	estate
planning	engineers
video	security

Notes: This table reports the most frequent words appearing in the business names of firms founded by intangible and tangible scientists, respectively. All company names are converted to lowercase, and common terms such as company, limited, etc., are excluded.

Table A14: Robustness Check: Effect of VC on Scientists' Entry with Logit and Poisson Models

Dependent Variable:	$\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}$		$\textbf{BusinessCount}_t \mid \textbf{NoBusiness}_{t-1}$	
Model:	Logit		Poisson	
	(1)	(2)	(3)	(4)
Post1979 × Intangible	0.1162**	0.1206**	0.1246*	0.1448**
	(0.0588)	(0.0598)	(0.0658)	(0.0659)
Intangible	0.1931***	0.2540***	0.1697***	0.2268***
	(0.0466)	(0.0483)	(0.0518)	(0.0547)
Post1979	0.5934***		0.6270***	
	(0.0421)		(0.0499)	
Constant	-7.203***		-7.161***	
	(0.0330)		(0.0398)	
Controls		Yes		Yes
Year FE		Yes		Yes
Observations	4,250,561	4,135,663	4,250,561	4,135,663
Pseudo R^2	0.00786	0.01392	0.00848	0.01482

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The dependent variable for Columns (1) and (2) is a binary indicator of whether a scientist started a business in a given year. The dependent variable for Columns (3) and (4) is the number of businesses started by a scientist in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. Control variables include indicators for the highest degree attained, gender, and birth cohort. Standard errors are clustered at the individual level. All specifications include year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A15: Robustness Check: Effect of VC on Scientists' Entry with Other Controls

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
Post1979 × Intangible	0.0292***	0.0383***	0.0292***	0.0195**
	(0.0074)	(0.0074)	(0.0074)	(0.0079)
VC Deals	0.0025***	0.0037***	0.0018***	
	(0.0005)	(0.0005)	(0.0005)	
Bank Branches	0.0091***		0.0059***	
	(0.0003)		(0.0003)	
Population		0.9939***	0.6061***	
		(0.0386)	(0.0323)	
Year FE	Yes	Yes	Yes	
Individual FE	Yes	Yes	Yes	Yes
County-Year FE				Yes
Observations	4,232,643	4,237,703	4,220,249	4,250,561
R^2	0.15547	0.15464	0.15564	0.15988

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *VC Deals* refers to the number of VC deals in the county-year. *Bank Branches* refers to the number of active bank branches in the county-year, based on FDIC data. *Population* is the total population in millions of the county-year. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A16: Robustness Check: Counties and VC Presence

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	Panel A: VC Counties			
	(1)	(2)	(3)	(4)
Constant	0.0840***	0.0840***		
	(0.0030)	(0.0030)		
Post1979	0.0786***	0.0786***		
	(0.0053)	(0.0053)		
Intangible	0.0168***	0.0168***	0.0168***	
	(0.0046)	(0.0046)	(0.0046)	
Post1979 × Intangible	0.0343***	0.0343***	0.0343***	0.0687***
	(0.0083)	(0.0083)	(0.0083)	(0.0091)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	183,092	183,092	183,092	183,092
R^2	0.00145	0.00146	0.00175	0.14056
	Panel B: Non VC Counties			
	(5)	(6)	(7)	(8)
Constant	0.0426***	0.0426***		
	(0.0039)	(0.0039)		
Post1979	-0.0004	-0.0004		
	(0.0056)	(0.0056)		
Intangible	0.0012	0.0012	0.0012	
	(0.0062)	(0.0062)	(0.0062)	
Post1979 × Intangible	0.0073	0.0073	0.0073	0.0114
	(0.0092)	(0.0092)	(0.0092)	(0.0092)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	1,379,975	1,379,975	1,379,975	1,379,975
R^2	0.00013	0.00015	0.00019	0.15466

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by private-sector scientists from 1970 to 1986. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. Panel A includes scientists living in counties with a VC presence, and Panel B includes scientists living in counties without a VC presence. VC presence is calculated as whether the county had any early-stage VC deals during the sample period. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A17: Robustness Check: Continuous Intangibility Scores

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
Post1979 × IntangibleScore	0.3903***	0.2955***	0.2950***	0.6010***
	(0.0689)	(0.0689)	(0.0689)	(0.0742)
Post1979 × TangibleScore	-0.0595	0.0009	0.0012	-0.0371
	(0.0536)	(0.0536)	(0.0536)	(0.0582)
IntangibleScore	0.1476***	0.1363***	0.1363***	
	(0.0383)	(0.0383)	(0.0384)	
TangibleScore	-0.0244	-0.0167	-0.0167	
	(0.0300)	(0.0299)	(0.0300)	
Post1979	-0.1845***	-0.1832***		
	(0.0558)	(0.0558)		
Constant	-0.0136	-0.0138		
	(0.0313)	(0.0313)		
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	7,688,461	7,688,461	7,688,461	7,688,461
R^2	0.00014	0.00041	0.00043	0.15309

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *TangibleScore* is a continuous variable indicating the cosine similarity between the work specialty of the scientist and the tangible specialty dictionary based on SciBERT embedding. *IntangibleScore* is a continuous variable indicating the cosine similarity between the work specialty of the scientist and the intangible specialty dictionary based on SciBERT embedding. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A18: Robustness Check: Removing Information Technology Related Scientists

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
Constant	0.0744***	0.0720***		
	(0.0025)	(0.0025)		
Post1979	0.0602***	0.0396***		
	(0.0043)	(0.0042)		
Intangible	0.0109***	0.0099**	0.0099**	
	(0.0039)	(0.0039)	(0.0039)	
Post1979 × Intangible	0.0218***	0.0138**	0.0137*	0.0350***
	(0.0070)	(0.0070)	(0.0070)	(0.0076)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	4,001,066	4,001,066	4,001,066	4,001,066
R^2	0.00012	0.00038	0.00040	0.15486

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986 by excluding the scientists working for Silicon Valley-related specialties (i.e., computer science, theory and practice of computation, computer non-numerical processing, and computer hardware design). The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A19: Robustness Check: Comparing California and Non-California Scientists

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	Panel A: California Scientists			
	(1)	(2)	(3)	(4)
Constant	0.2441***	0.2456***		
	(0.0218)	(0.0218)		
Post1979	0.3297***	0.3449***		
	(0.0420)	(0.0436)		
Intangible	0.0175	0.0181	0.0186	
	(0.0320)	(0.0320)	(0.0320)	
Post1979 × Intangible	0.3055***	0.3110***	0.3154***	0.4255***
	(0.0672)	(0.0678)	(0.0679)	(0.0775)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	183,092	183,092	183,092	183,092
R^2	0.00145	0.00146	0.00175	0.14056
	Panel B: Non-California Scientists			
	(5)	(6)	(7)	(8)
Constant	0.0582***	0.0575***		
	(0.0033)	(0.0033)		
Post1979	0.0406***	0.0352***		
	(0.0056)	(0.0055)		
Intangible	0.0356***	0.0348***	0.0349***	
	(0.0078)	(0.0078)	(0.0078)	
Post1979 × Intangible	0.0331**	0.0276**	0.0279**	0.0804***
	(0.0135)	(0.0135)	(0.0135)	(0.0142)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	1,379,975	1,379,975	1,379,975	1,379,975
R^2	0.00013	0.00015	0.00019	0.15466

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by private-sector scientists from 1970 to 1986. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. Panel A includes scientists living in California, and Panel B includes scientists living outside of California. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A20: Robustness Check: Removing Delaware from the Sample

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
	(1)	(2)	(3)	(4)
Constant	0.1033***	-0.0403***		
	(0.0034)	(0.0052)		
Post1979	0.1298***	0.1260***		
	(0.0058)	(0.0058)		
Intangible	0.0270***	-0.0110*	-0.0109*	
	(0.0056)	(0.0057)	(0.0057)	
Post1979 × Intangible	0.0772***	0.0734***	0.0734***	0.0460***
	(0.0101)	(0.0101)	(0.0101)	(0.0099)
Control		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	4,154,409	4,125,331	4,125,331	4,125,331
R^2	0.00042	0.00382	0.00388	0.13253

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986 by excluding the scientists residing in Delaware. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A21: Equality of Coefficients: Effect of Patent and Publication Activity on Business Formation

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$	
Model:	(1)	(2)
	Private Industry	University
Post1979 × Intangible	0.1929***	0.0088
	(0.0172)	(0.0132)
Post1979 × InventPre1979	0.0588***	0.2421***
	(0.0170)	(0.0817)
Post1979 × PubPre1979	0.1213*	0.0395
	(0.0673)	(0.0278)
Post1979 × Intangible3 × InventPre1979	0.5974***	0.8929***
	(0.1147)	(0.2262)
Post1979 × Intangible × PubPre1979	0.2046	0.0021
	(0.1477)	(0.0353)
Individual FE	Yes	Yes
Year FE	Yes	Yes
Observations	1,563,067	1,262,682
R^2	0.14862	0.15070

Notes: This table reports the OLS estimates of the effect of publication and patent activity on business formation. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *InventPre1979* equals one if the scientist is an inventor in a granted patent that was applied for before 1979. *PubPre1979* equals one if the scientist published a journal article before 1979. *Post1979* equals one for years after 1978. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A22: Heterogeneous Effect on Scientists with Propensity Score Matching

Dependent Variable:	$100 \cdot \mathbb{1}\left[\textbf{StartBusiness}_t \mid \textbf{NoBusiness}_{t-1}\right]$			
Panel A: Scientists Working in Private Industry or Sector				
	(1)	(2)	(3)	(4)
	Has Patent	No Patent	Has Publication	No Publication
Post1979 × Intangible	0.8038***	0.2069***	0.4084***	0.0594
	(0.1247)	(0.0521)	(0.1543)	(0.1569)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	246,917	229,757	26,256	28,067
R^2	0.15291	0.16540	0.15231	0.29895
Panel B: Scientists Working for Colleges or Universities				
	(5)	(6)	(7)	(8)
	Has Patent	No Patent	Has Publication	No Publication
Post1979 × Intangible	0.9016***	0.0196	0.0145	-0.0299
	(0.2426)	(0.0957)	(0.0349)	(0.0405)
Control	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	21,281	28,050	205,812	198,561
R^2	0.15386	0.28011	0.15880	0.19900

Notes: This table reports the difference-in-differences estimates of the ERISA effect on business formation by scientists from 1970 to 1986. Propensity score matching was implemented between inventors and non-inventors, and between publishing and non-publishing scientists, using nearest-neighbor matching (1:1 ratio) with covariates including the level of highest degree, gender, year of birth, and gross income. The dependent variable is a binary indicator of whether a scientist started a business in a given year. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on the LLM classification. *Post1979* equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A23: Effect of VC on the Industry Growth Rate

Panel A: Employment				
Dependent Variable:	$\log(\mathbf{Employee}_{t+1}/\mathbf{Employee}_t)$			
	(1)	(2)	(3)	(4)
	Full Sample		VC County	Non VC County
Post1979 × Intangible	0.0158***	0.0172***	0.0140***	0.0183***
	(0.0017)	(0.0017)	(0.0030)	(0.0021)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes			
County FE	Yes			
Year-County FE		Yes	Yes	Yes
Observations	376,524	376,524	83,873	292,651
R^2	0.02893	0.12503	0.06549	0.14817

Panel B: Establishments				
Dependent Variable:	$\log(\mathbf{EstablishmentCount}_{t+1}/\mathbf{EstablishmentCount}_t)$			
	(5)	(6)	(7)	(8)
	Full Sample		VC County	Non VC County
Post1979 × Intangible	0.0090***	0.0095***	0.0101***	0.0094***
	(0.0009)	(0.0009)	(0.0017)	(0.0010)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes			
County FE	Yes			
Year-County FE		Yes	Yes	Yes
Observations	840,011	840,011	123,347	716,664
R^2	0.02065	0.05581	0.03964	0.06056

Notes: This table reports the difference-in-differences estimates of the ERISA effect on the industry growth rate from 1974 to 1986. *VC County* represents a subsample with counties that had at least one VC investment during 1979-1986. *Non VC County* is the opposite. *Number of Employees* is the number of paid employees during the payroll period. *Establishments* is the number of establishments with paid employees. *Intangible* is a binary variable indicating whether the industry is classified as intangible. *Post1979* equals one for years after 1978. Standard errors are clustered at the county level. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table A24: Robustness Check: Intangible Capital and Fixed Asset Specificity

Panel A: Employment				
Dependent Variable:	**Number of Employees**			
	(1)	(2)	(3)	(4)
	Full Sample		VC County	Non VC County
Post1979 × RecoveryPPE	-2.074***	-2.659***	-17.47***	-0.7790***
	(0.2177)	(0.2634)	(2.019)	(0.0541)
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes			
County FE	Yes			
Year-County FE		Yes	Yes	Yes
Observations	1,377,151	1,377,151	179,205	1,197,946
R^2	0.38359	0.38797	0.42818	0.28559
Panel B: Establishments				
Dependent Variable:	**All Sizes**		**1-4 Employees**	
	(5)	(6)	(7)	(8)
	VC County	Non VC County	VC County	Non VC County
Post1979 × Intangible	-0.5672***	-0.0243***	-0.0231	0.0132***
	(0.0621)	(0.0027)	(0.0251)	(0.0014)
Industry FE	Yes	Yes	Yes	Yes
Year-County FE	Yes	Yes	Yes	Yes
Observations	179,205	1,377,151	179,205	1,197,946
R^2	0.41794	0.41983	0.37614	0.39069

Notes: This table reports the difference-in-differences estimates of the ERISA effect on industry size from 1974 to 1986. *VC County* represents a subsample with counties that had at least one VC investment during 1979-1986. *No VC County* is the opposite. *Number of Employees* is the number of paid employees during the payroll period. *Establishments* is the number of establishments with paid employees, categorized by employment-size class (e.g., 1–4 employees). *RecoveryPPE* is a continuous variable indicating the PPE liquidation recovery rate at the SIC two-digit level (Kermani and Ma, 2023). *Post1979* equals one for years after 1978. Standard errors are clustered at the county level. * $p < .10$, ** $p < .05$, *** $p < .01$.