Venture Capital and Scientists' Selection into Entrepreneurship

Xuelai Li*

March 16, 2025

Abstract

This paper examines the causal effect of venture capital (VC) on scientists' selection into entrepreneurship, using the 1979 clarification of the prudent-man rule under the Employee Retirement Income Security Act (ERISA) as a natural experiment. By relaxing pension fund allocation restrictions, the reform substantially expanded the pool of capital available to VC firms. I construct a novel historical dataset of US scientists in the 1960s, and track their business formation activities. I exploit the exogenous cross-sectional variation in how scientists' work specialties rely on tangible versus intangible assets. I show that the business formation rate of scientists doubled post ERISA, and the effects are stronger for those with intangible specialties, as they lacked bank funding due to insufficient collateral but are more attractive to VC. These scientists were not marginal entrants but productive inventors filing patents. I identify three potential mechanisms: (i) alleviating financial constraints, (ii) improving appropriability, and (iii) fostering localized entrepreneurial communities.

Keywords: Financial Intermediation, Venture Capital, Entrepreneurship, Innovation

*Finance Department, Imperial College Business School, x.li20@imperial.ac.uk. I am indebted to my supervisors, Ramana Nanda and Cláudia Custódio, for all their guidance. I also thank Jamie Coen, Rajkamal Iyer, Ziang Li, Lu Liu, Clara Martínez-Toledano, Tarun Ramadorai, Ailsa Röell, Magdalena Rola-Janicka, Savitar Sundaresan, Alex Whalley, Tong Yu; and seminar participants at the Wharton Innovation Doctoral Symposium and Imperial College London for their valuable comments and suggestions. All errors are my own.

1 Introduction

Venture capital (VC) has long been recognized as a driver of innovation and business formation (Nanda and Rhodes-Kropf, 2017; Howell, 2017). Macro-level evidence suggests that regions receiving greater VC inflows exhibit stronger economic growth (Samila and Sorenson, 2011). Highly skilled individuals forming technology entrepreneurship is especially important (Akcigit and Kerr, 2018; Christensen, 2011). Despite this recognized importance, identifying the causal impact of VC on the entrepreneurial decisions of these individuals remains a challenge.

The 1979 revision of the prudent-man standard under the Employee Retirement Income Security Act (ERISA) offers a setting to tackle this challenge. This ERISA reform 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 leveraging this policy change has yet to provide identification of VC's causal effects.

Another challenge is how to examine the selection into entrepreneurship decisions at the individual level. Usually, only those who start a business are observed. Moreover, because of the importance of technology entrepreneurship, we especially care about the behavior of high-skilled individuals, i.e., scientists. However, the decisions of scientists at the individual level are largely unexamined due to the lack of systematic data. Constructing a comprehensive dataset on these potential entrepreneurs is empirically demanding.

This paper addresses these gaps by causally estimating the impact of VC on scientists' selection into entrepreneurship, My empirical designs include two parts. First, I construct a novel panel of US scientists active in the 1960s by compiling a snapshot of their education backgrounds and work experiences, which I then link to business registration data to observe subsequent selection into entrepreneurship. Second, I exploit exogenous cross-sectional variation in scientists' work specialties by classifying the specialties according to reliance on tangible assets. VC seeks scalability and outsized returns, usually more common in less capital-intensive businesses. Therefore, scientists with intangible work specialties are more likely to be affected by the expansion of VC. More importantly, these individuals did not select their specialties in anticipation of future VC inflows, as the US VC market was negligible in the 1960s, and bank financing was generally more accessible for scientists with tangible specialties.

The main finding is that, following the ERISA reform, scientists became more likely to start businesses. Business formation by scientists more than doubled after the reform. Scientists with intangible specialties are 0.05% more responsive to the ERISA shock than those with tangible specialties. The effects are primarily driven by scientists working in the private sector and those who have filed a patent, with the effect size increasing to 1.17%. Given that the total business formation rate for all scientists is only 3.15%, the effect is non-negligible. I show that the main finding of a significant impact of VC on scientists' selection into entrepreneurship remains robust even when I (i) exclude computer science–related scientists, as they are strongly associated with Silicon Valley phenomenon, (ii) employ alternative tangibility-measure cutoff thresholds, or (iii) use continuous tangibility measures.

I further analyze three potential mechanisms behind the main effect. The first mechanism centers on financial constraints. Using intrastate branching deregulation as a negative credit shock to young firms (Chava, Oettl, Subramanian, and Subramanian, 2013; Hombert and Matray, 2017), I show that scientists engaged in tangible fields are more sensitive to disruptions in bank lending. However, those specializing in intangible areas typically do not rely on bank financing, so they are not affected by the deregulation. Instead, an expansion in VC availability alleviates the financial constraints of the scientists with intangible specialties that might otherwise not start a business. This suggests that VC serves as an important complement to conventional bank financing, which tends to favor more tangible and collateralizable projects.

The second mechanism involves appropriability. I show that scientists who have filed at least one patent and who are working in the private sectors are more responsive to the ERISA shock, i.e., spin out to commercialize their own research. This is because scientists working within private sectors may not fully capture the returns to their innovations, particularly when patents are assigned to the employer (Babina and Howell, 2024). VC, structured as an equity investor, does not encumber scientists' intellectual property.

The third mechanism is around entrepreneurial community building. VC tends to flow into regions with a history of government investment, creating localized entrepreneurial ecosystems. The results indicate that scientists residing in these counties are especially responsive to an influx of VC capital, as prior public funding has laid the groundwork for knowledge spillovers and network connections conducive to business formation (Nanda and Sørensen, 2010).

Overall, these findings show that VC significantly enhances the rate of business formation among scientists, particularly those with intangible working specialties. Far from being marginal entrants, these scientists often hold patents and are seen as productive inventors.

This paper contributes to three streams of literature. First, my results provide insight for the literature on financial intermediation and small business financing. Although prior research shows that small and medium-sized enterprises (SMEs) primarily rely on debt (Robb and Robinson, 2014) and home equity (Corradin and Popov, 2015; Kerr, Kerr, and Nanda, 2022), it also demonstrates that bank credit availability significantly influences the innovation activities of young firms (Chava et al., 2013; Hombert and Matray, 2017). Nevertheless, the evidence presented here indicates that banks fail to serve technology startups relying on intangible assets, leaving a funding gap that VC can bridge.

Second, this paper speaks to the venture capital 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). However, the mechanisms through which VC incentivizes high-skilled individuals to start a business are not yet fully understood. The paper shows three mechanisms: financial constraints, appropriability, and entrepreneurial community. Moreover, I show that once VC supply expands, the individuals that select into entrepreneurship are not marginal entrants but are productive inventors.

Finally, the findings link literature between financial intermediation and government expenditure, traditionally centered on bank lending and fiscal multipliers (Goldman, Iyer, and Nanda, 2022). While earlier studies emphasize public R&D crowds in private investments in the mid-term (Antolin-Diaz and Surico, 2022) and generates spillovers to large-firm R&D (Azoulay, Graff Zivin, Li, and Sampat, 2019; Moretti, Steinwender, and Van Reenen, 2023), recent research suggests that government-funded R&D can catalyze private capital investment by de-risking nascent technologies (Rezaei and Yao, 2024). This paper offers new evidence on how financial intermediaries complement public R&D by supporting scientists' business formation. Specifically, VC investment selectively targets government-invested industries and locations. Public R&D expenditures both mitigate technological risks and promote human capital formation through on-the-job training, thereby attracting VC participation. In turn, this VC inflow releases the entrepreneurial potential accumulated under government-funded research.

The rest of this paper is organized as follows. Section 2 provides an overview of the historical context of financial intermediaries for 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 VC supply and business formation, distinguishing by the tangibility of scientists' specialties. Section 5 explores three mechanisms. Finally, Section 6 concludes.

2 Historical Context

2.1 Emergence of Risk Capital Intermediation

The financing landscape for technology entrepreneurship remained largely informal until the advent of venture capital in 1959, marked by the establishment of Draper, Gaither & Anderson (DGA), the first Silicon Valley venture capital 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 SBICs, were off-limits to those unwilling to accept government loans, which many perceived as restrictive to growth-oriented firms. 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 venture capital. This left individual investors as a potential source of funding; however, 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 venture capital 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 and regulatory constraints on pension fund investments further restricted the growth of the venture capital industry, leaving early-stage startups with limited funding opportunities. Before the ERISA reform in 1979, the "prudent man" rule made many pension managers not dare to put money into VC funds, as investing in small business securities can be seen as imprudent. The ERISA uniformed the fiduciary requirement of private pension funds. A fiduciary must discharge its duty "with the care, skill, prudence, and diligence under the circumstances then prevailing that a prudent man acting in alike 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, especially the "prudent man" rule, make many pension managers not dare to put money into VC funds, as investing in small business securities can be of high 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 to the Congress, which was approved in October, on the reorganization plan.[1] The Treasury will have statutory authority for minimum standards, while the Department of Labor (DOL) will have statutory authority for fiduciary obligations.

In June 1979, the DOL explicitly clarified the fiduciary requirement in a federal register (details in Figure A1), allowing fund managers to invest their capital in venture funds. This reform significantly increases the supply of capital to VC funds, as shown in Figure A2. The fundraising patterns are mirrored in the investments by venture capitalists into small firms (Kortum and Lerner, 2000).

The composition of limited partners in VC funds changed significantly due to the ERISA reform. Pre-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

[1] According to message to the Congress Transmitting Reorganization Plan No. 4 of 1978.

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.[2] While ERISA exclusively affects private pension funds, these funds generally exhibit greater allocations to VC compared to state pension funds.

2.2 Bank Intermediation in Financing SMEs

During the 1970s, commercial banks in the US were the primary source of financing for SMEs. Their dominance largely stemmed from the Glass-Steagall Act, which confined banks to traditional commercial banking activities and prevented them from engaging in securities trading or insurance. This legal environment, coupled with restrictions on interstate banking and branching, left most banks focused on local economies through deposits, loans, and savings accounts. In turn, the highly fragmented banking sector, composed of numerous small and regional banks, limited competition and risk diversification. Furthermore, high interest rates made it more challenging for startups to secure bank loans. During the 1980s, when US interest rates peaked at 21.5% as the Federal Reserve attempted to curb inflation, debt financing became particularly costly.

Although certain regulatory adjustments began taking shape around 1980, they did not align with the ERISA reforms. For instance, the Depository Institutions Deregulation and Monetary Control Act of 1980 eased restrictions on deposit interest rates, while the Garn–St. Germain Depository Institutions Act of 1982 granted savings institutions the ability to engage in riskier lending activities.

The banking sector underwent two major deregulatory shifts after the 1970s: interstate and intrastate deregulation. The existing literature on *interstate* banking deregulation suggests that it allowed banks to achieve greater geographic diversification, thereby expanding credit supply to innovative firms (Cornaggia, Mao, Tian, and Wolfe, 2015). However, states did not begin relaxing restrictions on interstate banking until the 1980s, which falls outside my sample period.

[2] The largest state pension funds (e.g., CalPERS, CalSTRS, NYSCRF, Texas TRS) have some VC exposure but allocate a relatively small proportion of their total assets to VC compared to private pensions.

In contrast, *intrastate* deregulation represents a negative credit shock to SMEs. Before 1970, most states either prohibited or strictly curtailed branching. Starting in 1970, the remaining states gradually lifted these barriers through a three-stage process. First, they allowed multibank holding companies to form; next, they enabled branching through mergers and acquisitions (M&A) only; and ultimately, they sanctioned unrestricted (de novo) branching. This evolution facilitated broader geographic expansion and encouraged greater competition (Jayaratne and Strahan, 1996; Black and Strahan, 2002). Empirical studies document a decline in innovation following intrastate deregulation, attributing this effect to two potential mechanisms: (1) increased market concentration reduces credit availability for SMEs (Chava et al., 2013), or (2) heightened banking competition that weakens relationship lending (Hombert and Matray, 2017).

3 Historical Data

3.1 Scientists and Engineers

This paper investigates the supply of VC and scientists' selection into entrepreneurship. The ERISA reform happened in 1979, so I collected a list of scientists active in the 1960s and tracked their business formation activities from 1970 onward. To comprehensively understand the US technical personnel in 1960s, I collected individual-level data from two sources: the National Register of Scientific and Technical Personnel from the National Archives and the American Men of Science.

3.1.1 National Register of Scientific and Technical Personnel

I retrieved the National Register of Scientific and Technical Personnel (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 evolved into a key source of statistical information on scientific and engineering personnel.[3] It provides critical data for developing national

[3]https://aad.archives.gov/aad/series-description.jsp?s=3550

science policy and supplies information to Congress and government agencies.

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 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 in 8 years.[4] The register was originally established to identify specialized personnel during national emergencies. However, their utility for statistical analysis was soon recognized, leading to a shift in their primary function towards providing statistical information for scientific and engineering personnel. The content of each record varies slightly year by year, but typical entries include details such as name, institution, sex, age, educational background, employment specialty, job function, 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 paper uses the 1962–1968 NRSTP data because these four waves include the scientists' residence city information.

The series was disseminated through various academic societies. Respondents were predominantly academic and research professionals. While the content varies annually, each record typically contains information on the individual's name, demographical information, educational background, employment specialty, and self-reported income. This dataset serves as a comprehensive source for understanding the workforce during these periods. The response rate is approximately 60% but varies across academic societies. For instance, in 1968, the response rate among biologists was 54%[5], while around 70% of eligible individuals were included in the Register of the American

[4] 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.

[5] American Institute of Biological Sciences Annual Report 1969

Meteorological Society [6]. Additionally, the NSF reported that over 90% of US science doctorates were captured in the 1964 wave of the survey.

The data was processed by extracting information from the digitized codes, as shown in Figure 1. Subsequently, the codes for each variable were matched with their meaning, which was documented in the photocopies of codebook films. The raw digitized format consists of thousands of entries, with each line representing an individual record. The values in different positions in the line correspond to different variables (i.e., survey questions). To analyze the data, I first separate these values into their respective variables. Subsequently, I matched numbers with their descriptions based on codebooks. The codebooks were scanned documents without optical character recognition (OCR). So, I manually cleaned the codebooks to ensure accurate mapping between numerical values and descriptions. When the original scan is faint, certain words are best guesses based on common nomenclature by ChatGPT-4o. ChatGPT excels at this task, as transformer models are trained to reconstruct incomplete sentences and words.

3.1.2 American Men of Science

I digitized the Eleventh Edition of American Men of Science (AMS) that was collected from 1960 to 1965. AMS is a comprehensive directory of scientists across the United States and Canada. First published in 1906 by James McKeen Cattell, the AMS is an exceptionally comprehensive source of biographical information for male and female scientists. Cattell collected these data originally for his own research on the psychology of intelligence. Born into a wealthy Pennsylvania family, Cattell earned his PhD in Leipzig, Germany, and became the first American to publish a dissertation in psychology (Airoldi and Moser, 2024).

The inclusion criteria are based on scientific achievement, research quality, and responsibility in science-based positions. This edition was compiled at Arizona State University and follows a pattern similar to the Tenth Edition, with efforts to keep biographical information up to date through questionnaires. The project has a long his-

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

tory, originating in 1906 and growing into a major reference resource for American science.

AMS was created 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. The criteria for inclusion in AMS are, as per the Preface to this edition

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 split into two sections: Physical and Biological Sciences and Social and Behavioral Sciences. Only the first section was digitized as the primary focus of this research is on the scientific and technical personnel. It contains around 21,000 biographies per volume, with a total of over 150,000 entries.

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

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

3.1.3 Concatenating the Two Data Sources

I first dropped all the scientists whose county location information was missing because the later matching process relied on both name and location. The NRSTP dataset has 10% of county FIPS missing, while AMS has 45% missing[7].

After the data processing, my NRSTP sample records include 447,317 scientists who responded to the survey between 1962 and 1968. The AMS includes 59,877 scientists who appeared in the 1965 edition. 31,468 scientists appear in both datasets based on name and county location information. For the overlapping entries, I retained the records in the NRSTP because the variables recorded in NRSTP are more comprehensive than those in AMS. 56% of the AMS scientists appear in the NRSTP records, indicating that NRSTP has a good record of senior scientists. Thus, AMS serves as a complementary dataset to the NRSTP records on the senior scientists. I dropped 392 scientists whose work specialties are not correctly recorded. The final dataset contains information on 475,334 scientists.

3.2 Business Registration

Business registration in the US is stored separately by each state's Secretary of State. OpenCorporates gathers the data and distributes it as a one-off download package.[8] This study used data from all jurisdictions (i.e., states) within the US. It is worth pointing out that bankruptcy or any other type of litigation against the company is not in the records of the Secretary of State. Instead, this type of information would have to be

[7]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 US Department of Housing and Urban Development's USPS ZIP Code Crosswalk Files.

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

discovered through a litigation search.[9]

The business registry data from OpenCorporates covers 76 million businesses across all US states. The data includes incorporation dates and dissolution dates. The data indicate the state and registration address for the business. Most businesses are registered in the same state listed as their address, but businesses can also 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). Also, many firms are registered in the state they operate in and Delaware. OpenCorporates covers both and often connects the two with 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 provides business registry data dating back to the 1940s or earlier, depending on the state's records. It includes officers' names linked to companies, which is essential for matching with the scientists' data. Therefore, OpenCorporates provides the a consistent publicly available dataset on US business registrations.

Business addresses in OpenCorporates are cleaned by using regularization to extract the zip codes and then match them to the corresponding county. During the period which I used to match with scientists (1945-1990), 57% of the 14,495,168 firms in the dataset possess complete registered address data. Among these firms, 84% include zip code information. Utilizing the OpenStreetMap API, I enhanced the coverage of zip code information to 93% by extracting zip codes from the remaining 16% of non-standard addresses. This process adds zip code information for 155,820 addresses through the API.

[9]https://www.jonesday.com/en/insights/2012/10/public-disclosure-requirements-for-private-companies-us-vs-europe

3.3 Matching Data

3.3.1 Matching Scientists and Engineers with Business Registration

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

Then I map the OpenCorporates data with the AMS and NRSTP data by name and county FIPS code. I only matched 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 individuals with the same name but different identities becomes more significant. My final data consists of 28,075 firms matched to 14,967 scientists with at least one firm recorded. 3.15% of the scientists are found to be associated with at least one business.

3.3.2 Matching Scientists and Engineers with Patent Data

Patent data is from the PatentCity dataset (Bergeaud and Verluise, 2024), which provides information on the zip code and inventor names of US patents back to 1836. Compared 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.

Then I map the Patent data with the AMS and NRSTP data by name and county fips code. Again, I only matched scientists to businesses filed between 1900 and 2000 to reduce the risk of mistakenly matching individuals with the same name but different identities. My final data consists of 184,362 patents matched to 42,990 scientists with at least one patent recorded. 9.04% of the scientists are found to be associated with at least one patent.

3.3.3 Matching Scientists and Engineers with Publication Data

To measure scientific productivity, I match 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 scientists and engineers with author_ids in the MAG, using first and last names, as well as county fips of the affiliation of the paper author. Based on the birth year information of the scientists and engineers data, I further restricted matched publications to those who no longer published anything after 2005. My final data consists of 1,010,217 publications matched to 45,019 scientists with at least one publication recorded. 9.47% of the scientists are found to be associated with at least one paper publication.

3.4 Descriptive Statistics

My final sample consists of 475,334 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 supplemented 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 checked the NRSTP for gender information and used it if available. If not, I applied the gender guesser to predict the gender based on the scientist's first name. The scientists and engineers sample is dominated by males, with 433,451 male scientists and 39,903 female scientists. This is consistent with the literature.

Cohort The NRSTP dataset does not have Date of Birth information as the AMS, so I developed a method to predict the year of birth of scientists based on the Year of Highest Degree and the Level of Highest Degree recorded in the NRSTP. I assumed that individuals 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 estimated the year of birth by subtracting the predicted age at the time of obtaining the highest degree from the Year of the Highest Degree, 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, shaping a more risk-averse and pragmatic outlook (Figure A3).

Education The complete data sample comprises 454,383 scientists, including 167,700 PhD holders and 10,973 MD holders. The average year in which 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 scientists are the University of Michigan-Ann Arbor, Columbia University, and the University of California-Berkeley. Additionally, elite institutions such as Harvard University and MIT are also among the most common alma maters (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 US.[10] This shows that the core technical personnel in the US are significantly more centralized. This concentration suggests that the critical expertise and resources were likely pooled in specific

[10]For example, during the Cold War, Natrona County (FIPS 56025), Wyoming, was involved in uranium mining, which was crucial for nuclear weapons development. El Paso County (FIPS 08041), 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.

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 Scientists earned more than the general population at the lower and middle quantiles (Table A3). The inequality within the scientific community is less than the overall US population. These reflect the specialized skills, advanced education, and relatively standardized wage structures within scientific professions.

Employment Most 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 from chemical manufacturing, petroleum-related industries, electrical and electronics sectors, and large aerospace and defense contractors A5).

Work Specialty A major challenge was to compile work specialties into an individual-year panel. The NRSTP generated a sequence of identifiers for each specialty in each wave of the survey. However, these identifiers varied 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 specialties across years, I standardized names and manually merged or split the specialties as needed. The data sample reveals a strong educational background concentration in Chemistry, Biology, and Geology (Table A6). This aligns with the fact that the top private employers of scientists are primarily companies in the chemical manufacturing and petroleum and coal products manufacturing industries (Table A5). Table A7 further illustrates top specialties that are funded by the government defense or space programs.

Patent and Publication 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 have been cited by approximately 30,000 patents. There is a weak correlation between business formation activity and both patenting and publication activity, indicating a limited association between these factors (Table A8). This suggests that scientific output and intellectual property generation do not strongly predict entrepreneurial activity among scientists.

4 Main Results

I use the 1979 ERISA reform as an exogenous shock that led to the large-scale emergence of VCs as a financial intermediary. First, this reform is unique in its significant impact on VC fundraising, as one of the few regulatory changes to do so. While the capital gains tax cut in the 1980s could also influence 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, Lerner, Blair, and Hellmann, 1998). Second, the early-stage equity investment landscape of the 1980s did not have a standardized approach yet. Equity investment in small businesses was primarily provided by individuals, with little involvement from financial intermediaries. Moreover, 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. Post 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). Between 1970 and 1978, 4,149 scientists started a business, but this number surged to 7,936 during 1979–1986. Business formation by scientists more than doubled following the ERISA reform, indicating its unique impact on scientists. Notably, the total business formation rate for scientists, based on business registrations from 1945 to 1990, is 3.15%.

4.1 Measure of the tangibility of specialties

The 1979 ERISA reform is a one-off exogenous shock. I leverage the scientists' work specialties to create cross-sectional variation for a difference in differences design. Scientists working in fields more reliant on intangible assets likely face greater exposure to the ERISA shock, given private capital's tendency to invest in less capital-intensive industries. As shown in Figure A5, the investment theme of the venture capital industry has transformed from computer hardware and electronics to less capital-intensive but high-growth-potential industries such as business services.

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

To distinguish between tangible and intangible work specialties, I utilize a large language model with the dictionary method (Ash and Hansen, 2023). I first used ChatGPT-4o to create two dictionaries: one for tangible specialties and one for intangible specialties. The contents of these dictionaries are listed in Table A9. 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 and a word by capturing semantic relationships in a continuous vector space. Unlike bag-of-words, which relies on word frequency and ignores context, embeddings account for meaning and word associations, enabling more accurate comparisons (Li, Mai, Shen, and Yan, 2021). This is particularly valuable in my context, as it allows for handling 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.

After calculating the similarity scores, I compute the absolute difference between the scores for intangible and tangible categories. The distribution of score differences between tangible and intangible similarity is presented in Figure A6. Some specialties exhibit similarity to both the tangible and intangible dictionaries, either closely or distantly. For instance, exfoliative cytopathology[11] has both low tangible and intangible similarity scores, with minimal difference between them. This suggests that textual similarity alone does not clearly categorize this specialty as either tangible or intangible. A specialty is classified as tangible or intangible if the absolute difference in scores exceeds 0.04, with classification determined by the higher score. If the absolute difference is 0.04 or less, no classification is assigned.

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

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.

[11] A branch of cytopathology that involves the study of cells shed or scraped from epithelial surfaces or body fluids to diagnose diseases, including infections, inflammatory conditions, and cancers.

4.2 Identification

Below is the linear probability model with a Difference-in-Differences (DiD) estimator.

$$Y_{it} = \beta Intangible_i * Post1979_t + X_{ct} + \eta_i + \eta_t + \epsilon_{it} \tag{1}$$

Y_{it} represents the outcome variables, including business formation and patenting activities. $Intangible_i$ 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. X_{ct} are county-year level control variables.

The ideal experiment would randomly assign scientists with varying exposure to VC availability. I use the tangibility of scientists' work specialties in the 1960s as a cross-sectional variation in their exposure to VC. The key assumption is that VC investors are more inclined to finance intangible businesses due to their scalability and high return potential, whereas banks primarily fund tangible businesses backed by collateral. The work specialties of scientists in the 1960s can be considered exogenous, as VC was almost nonexistent at the time, and scientists did not select their fields based on the potential to launch intangible businesses. In fact, scientists specializing in tangible fields had an advantage in starting businesses, as they could secure bank financing using collateral such as machinery. To identify the effect of VC supply, I exploit ERISA as an exogenous shock and compare business formation rates between scientists with tangible and intangible specialties, where the latter group is more affected by changes in VC availability.

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 controls in column (1), year fixed effects in column (2), and also scientist individual fixed effects in column (4). The dynamic specification in Figure 4 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.

The results in Table 4 also show that following the 1979 ERISA reform, scientists with more intangible work specialties are significantly more likely to file patents. In contrast, their likelihood of publishing academic papers declined. This pattern is consistent with the notion that publishing is less aligned with the commercialization of technology compared to patenting. The results are robust by adding controls, year fixed effects in Columns (1) and (3), and also scientist individual fixed effects in Columns (2) and (4).

I also conduct robustness checks using different definitions and subsamples, demonstrating that the effect remains robust. Table A13 and Table A14 show 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. Table A15 further confirms that using a threshold of 0—without imposing a 0.04 difference between the tangibility and intangibility scores to classify Intangible—yields consistent results.

Overall, the results indicate that intangible scientists are 0.05% more likely to start a business than tangible scientists. Given the overall business formation rate of scientists at 0.8% before the ERISA shock, this corresponds to a relative effect of approximately 6%. Moreover, the patenting results show that intangible scientists are 0.16% more likely to start a business than tangible scientists. Given the overall business formation rate of scientists at 9%, this translates to a relative effect of approximately 1.8%. This evidence suggests that the influx of private capital effectively alleviates financial constraints faced by scientists, thereby fostering innovation and entrepreneurial activity.

5 Mechanisms

5.1 Financial Constraints

According to the literature, one hypothesis is that banks may also finance intangible assets as young firms rely on external debt (Robb and Robinson, 2014). If this holds, then simply improving access to banking credit could also stimulate business formation among scientists. In a perfect capital market, firms are indifferent between financing through debt or equity.

To test this counterargument, I should examine whether the banking sector increases (or decreases) its financing of intangible assets in response to changes in credit supply. Intrabank deregulation occurred during the sample period, which decreased the banking credit availability to small firms. During the 1970s and 1980s, the banking industry underwent consolidation. Beginning in the early 1970s, 35 states implemented deregulation measures that relaxed restrictions on intrastate branching, allowing bank holding companies to consolidate subsidiaries into branches and permitting statewide de novo branching (Jayaratne and Strahan, 1996). Black and Strahan (2002) finds that the rate of new business incorporations increased following these deregulation efforts, as banks appeared to lend more effectively. However, Chava et al. (2013) find that although intrastate deregulation created more efficient banks, it also increased banks' bargaining power over small firms. Compared to small banks, large banks lend disproportionately less to small firms.

Following the line of intrabank deregulation research, I employed an event study design to examine whether the deregulation of intrastate branching restrictions led to increased business formation in states that adopted these policies. The lifting of branching restrictions in the US banking sector occurred through two key channels: M&A and de novo branching. Before deregulation, strict interstate banking laws prevented banks from expanding across state lines. Deregulation allowed banks to expand via M&A, enabling larger institutions to acquire existing banks in other states and immediately integrate their branch networks. Alternatively, de novo branching allowed banks to open new branches from scratch in previously restricted areas.

$$BusinessFormation_{ist} = \beta Deregulation_{st} + \eta_i + \eta_t + \epsilon_{ist} \tag{2}$$

$Deregulation_{st}$ 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.

I used the Callaway and Sant'Anna DiD estimator to estimate the dynamic treatment effect, which accounts for treatment heterogeneity by comparing treated units to never-treated units (Callaway and Sant'Anna, 2021). Figure 5 shows that lifting M&A restrictions in scientists' states negatively affects business formation. There is no evidence of differential pre-trends, suggesting that the parallel trends assumption holds. The two subfigures show that, post-deregulation, scientists with tangible specialties are more affected by the shock, whereas scientists with intangible specialties are not significantly impacted.

The event study on intrastate banking deregulation shows that, although lending efficiency increases post-deregulation, credit to small firms declines, leading to lower business formation among scientists, especially those whose specialty is more tangible. This is consistent with the literature, which finds that young firms file fewer patents after interstate deregulation (Chava et al., 2013; Hombert and Matray, 2017). Note that the development of high-yield bonds, pioneered by Michael Milken, allowed smaller or riskier firms to access the bond market, yet these instruments primarily targeted businesses with tangible assets that could serve as collateral.[12] In contrast, VC not only provides capital but also offers startups critical managerial expertise, strategic guidance, and valuable industry connections, further distinguishing it from conventional financing sources.

[12]Notably, one-fifth of the bonds issued between 1978 and 1983 had defaulted by 1988 (Greenspan and Wooldridge, 2018).

Through the event study using intrastate deregulation as a negative bank credit availability shock, I demonstrate that VC represents a distinct form of capital that complements rather than substitutes for traditional banking institutions by financing the intangible scientists. VC exhibits a unique risk appetite and investment preference, with a strong emphasis on intangible, innovation-driven businesses that are traditionally underserved by the banking sector.

5.2 Entrepreneurial Spawning

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 firms 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 VC supply. Indeed, Figure 6 shows that university scientists are not responsive to the increase of VC supply, whereas the effect is significant to scientists employed in the private sector.

Table 5 indicates that the business formation effect is primarily driven by private sectors, instead of those who are working in the university, federal governments, or self-employed. This indicates a spinout effect. Entrepreneurial spawning happens when individuals become entrepreneurs because the large bureaucratic companies for which they work are reluctant to fund their entrepreneurial ideas (Gompers, Lerner, and Scharfstein, 2005). Employees of large firms thus create spin-out businesses by leveraging their experience and expertise. A widely used 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 copier business. Most of the value from Xerox's inventions was captured by employees who left to start companies like Adobe and 3Com.

Engineers employed in large firms may be motivated to transfer technology through business formation, appropriating their expertise.[13] Yet, the lack of financing for the transition across the valley of death makes potential startups fail to get started.[14] Moreover, though possessing technical knowledge, engineers may lack the business acumen and network connection essential for entrepreneurship.[15]

I examine what kind of scientists are spinning out from the private sector. Does VC encourage high-quality startups, or does it primarily facilitate spinouts from low seniority and less productive scientists who cannot earn sufficiently high wages, so they decide to spin out? First, I also cross-validate the scientists' salary in my data with the National Aeronautics and Space Administration (NASA) historical salary scheme based on the self-reported salary in the survey. Figure A7 illustrates the density distribution of self-reported base salary. Most of the scientists and engineers are receiving salaries comparable to the 11- out of 18-grade salary rates at NASA. Note that grade 18 is the highest rate of salary, 11-grade is likely associated with a middle management level. This means that scientists in my sample are usually not in the top management team.

Furthermore, I examine the productivity of the scientists and their business formation decisions. In Table 6 Panel A, Columns (1) and (2), show that scientists who have filed at least one patent and are employed in the private sector are significantly more likely to spin out. The effect is approximately 1%, indicating a 30% increase in the likelihood of starting a business with a baseline rate of 3.15%. This substantial effect aligns with the argument that inventors seek to appropriate value from their inventions, but large firms often capture most of the benefits, creating an incentive for them to spin out. Columns (3) and (4) show that scientists who have published a critical journal

[13] Scientists and engineers who made major discoveries to their employers may get only token rewards. An ironic example is the $2 compensation from Raytheon to Percy Spencer for his invention of the microwave oven in 1945.

[14] 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.

[15] As venture capital 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).

article are 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.

Since many scientists work in universities, it is also worth examining whether university scientists' patenting and publishing activity are related to business formation decisions. University scientists may face higher costs when starting a business due to regulatory expenses and inherent preferences. In Table 6, Panel B, Columns (5) and (6) show that university scientists who have filed at least one patent are significantly more likely to start a business. However, the share of university scientists who file patents is low. There is no significant effect of business formation among university scientists who publish journal articles, as shown in Columns (7) and (8). This differs from the results in Panel A, which indicate that university scientists are less likely to start a business compared to industry scientists and that publishing papers does not facilitate business formation. Instead, filing patents appears to be beneficial for starting a business and may be linked to preferences for business formation.

5.3 Entrepreneurial Community

In the previous sections, I demonstrated that VC incentivizes scientists to start a business through alleviating financial constraints and appropriating their own innovation. However, literature has shown a broader effect by VC as it can spillover through coworker network (Nanda and Sørensen, 2010) and local community so generate affregated effects (Samila and Sorenson, 2011). This section explores whether VC incentivizes scientists to start a business by fostering an entrepreneurial community in their place of residence.

First, VC may flow into regions that already possess ex-ante suitability for business formation. I hypothesize that areas historically funded by the government, long before the ERISA shock, may have been particularly attractive to VC investment. These selected regions should be exogenous, as government funding decisions were not made to optimize business formation but rather to support strategic national programs.

Second, scientists residing in these regions may subsequently experience a greater impact from VC investment, as the influx of VC capital fosters the development of an entrepreneurial community in their local area.

To test this hypothesis, I leverage a significant exogenous shock to R&D investment in the 20th century: the *Space Race* between the United States and the Soviet Union. The Space Race concluded just seven years prior to the ERISA reform, making it a relevant historical event for analyzing long-term government investment effects. Figure 7 further illustrates the potential connection between ERISA and the Space Race, showing that business formation increased most post-ERISA among scientists who were previously funded by defense and space programs.

The Space Race led to an R&D windfall beginning in Fiscal Year 1962, marked by a significant increase in NASA's budget following deliberations between Congress and the federal government. As illustrated in Figure A8, NASA received an average of 2.5% of the federal budget during this period, peaking above 4% in 1964 and 1965. The Apollo Program alone accounted for nearly $30 billion in expenditures. The effective conclusion of the Space Race in 1972, with the final Apollo 17 mission, was followed by a relative decline in NASA's funding, which has since ranged between 1% and 0.4% of total federal spending.

The R&D investments made during the 1962–1972 period resulted in significant human capital accumulation. More importantly, the philosophy of NASA's contracting leads to human capital accumulation in the private industry. NASA separates the evaluation and production by delegating the technical direction and monitoring to the centers.[16] NASA itself does not set up production capacity that already exists in the private sector. As shown in Figure A9, there was a significant expansion in NASA's workforce, with the number of civil servants increasing from 10,200 at the beginning of Kennedy's presidency to 34,500 by the end of 1965. Additionally, the contractor workforce associated with NASA experienced even more rapid growth, reaching a total of

[16]Levine, Arnold S. Managing NASA in the Apollo era. No. 4102. Scientific and Technical Information Branch, National Aeronautics and Space Administration, 1982.

376,700 by the end of 1965.

5.3.1 Flow of Venture Capital

The first hypothesis posits that VC flows into regions that received significant government investment during the Space Race. However, the geographic distribution of these investments can be considered exogenous, as the Space Race was driven primarily by strategic and geopolitical objectives rather than commercial or civilian applications.

I investigated the flow of VC from both geographical location and industry focus. The geographical location reflects whether private capital flows to where there are talents, whereas industry focus shows whether the federally funded technologies are more favorable by private investors. To address the first dimension, I use the industry-county level space score from Kantor and Whalley (2024) to measure *SpaceCountyIndustry*. The construction of the space capability measure relies on digitized National Intelligence Estimates (NIE) on Soviet space technologies from 1947 to 1991. County-industry pairs are classified based on their pre-Space Race spaceflight technology by identifying similar technologies in US patents before Sputnik's launch in 1957. Using text similarity methods based on term frequency, the approach estimates the technological proximity between NIE documents and US patents. Aggregating these similarity scores across pre-1958 patents within county-industry cells forms the final space capability measure.

I conduct a DiD analysis based on the specification below.

$$Deals_{ict} = \beta_0 + \beta_1 SpaceCountyIndustry_{ic} * Post1979_t + \theta_i + \gamma_{ct} + \epsilon_{ict} \quad (3)$$

$Deals_{ict}$ is the number of VC deals or total amount of VC investment in the certain county-industry. $SpaceCountyIndustry_{ic}$ is exogenous because it simply reflects the technology and places that are suitable for the space program as defined by the Soviet Union. Consequently, it does not capture any potential selection bias associated with NASA's actual investment decisions.

Results in Table 7 show that following ERISA deregulation, private capital investment increasingly flows into county-industries that were more likely to have received Space Race investments. Columns (1) and (2) report the total log deal size of early-stage VC investments in a given county-industry, while Columns (3) and (4) present the number of deals. Columns (1) and (3) include year, industry, and county fixed effects, whereas Columns (2) and (4) incorporate year and county-year fixed effects. All interaction terms are statistically significant, suggesting that VC investment is concentrated in regions rich in scientific and technical talent. However, it is worth noting that, in the absence of the Post1979 term, space-related industries and counties received relatively little VC investment. This is likely due to the fact that, prior to ERISA, venture capital activity was limited, and many technologies developed during the Space Race were subject to non-disclosure agreements (NDAs) and other confidentiality restrictions. These limitations may have hindered the commercialization of such technologies, thereby reducing investment opportunities for VC.

These results validate the hypothesis that a significant number of VCs invest in places that were more likely to receive public R&D investments during the Space Race.

5.3.2 Business Formation of Scientists in Space Counties

In this section, I investigate whether scientists living in the space counties exhibit a higher propensity to start a business because of the entrepreneurial community fostered by the VC. I utilize the *SpaceCounty* definition from Kantor and Whalley (2024) and classify scientists residing in counties with an above-median space score as treated. These counties were more likely to receive NASA funding, as they had already produced innovation related to the Space Race before 1958.

Table 8 presents the regression results based on whether scientists reside in Space Counties and whether they are employed in the private sector. Columns (1) and (3) show that scientists living in Space Counties are more responsive to the VC shock, suggesting that VC investment in these regions fosters an entrepreneurial community that attracts scientists to start businesses.

Moreover, the effect on business formation is particularly strong for scientists who both reside in Space Counties and work in the private sector, as shown in Column (2). This finding is consistent with the historical context of the Space Race, during which NASA opted not to develop in-house production capabilities but instead relied on outsourcing to industry partners for technology development and manufacturing. By collaborating with a network of contractors, NASA leveraged external expertise to achieve its mission objectives. This procurement model benefited industrial firms by funding facilities, workforce recruitment, and technical training. As a result, scientists working in the private sector in Space Counties likely accumulated higher levels of human capital, increasing their propensity to spin out and form new ventures. Figure 8 further illustrates the results presented in Columns (2) and (4), confirming the findings. Notably, there is no evidence of pre-trends, supporting the validity of the parallel trends assumption.

6 Conclusion

This paper contributes to the growing literature on technology entrepreneurship by demonstrating that an expanded supply of VC can induce scientists to create businesses. Exploiting the 1979 ERISA shock as a quasi-exogenous increase in the availability of VC, I show that the rate of business formation among scientists doubled, especially significant among scientists with intangible specialties. Moreover, these scientists are not marginal entrants, they are productive inventors who have filed patents.

There are three potential mechanisms behind the main results. First, VC alleviates financing constraints for scientists whose projects lack collateral and traditional banks are typically unwilling to invest. In line with this view, I show that negative credit shock triggered by intrastate bank deregulation reduces business formation only among scientists with tangible specialties; there is no observable decline among those with intangible specialties. Second, VC encourages scientists in the private sector to spin out and appropriate their knowledge, particularly those who hold patents. This evidence underscores the role of equity financing in incentivizing inventors to commercialize innovations outside of corporate boundaries. Third, VC firms selectively invest into industries and locations that have benefited from prior public R&D spending, helping to cultivate entrepreneurial communities in these regions. This capital inflow unlocks the entrepreneurial potential accumulated under public R&D investments. This shows the critical role of VC in translating publicly supported scientific advances into commercial applications.

This paper offers clear policy implications. In particular, the ongoing debate over increasing pension fund allocations to VC by the United Kingdom[17] resonates with the evidence presented here. 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. VC plays an important role as a specialized financial intermediary for incentivizing technology entrepreneurship. Compared to government grants (e.g., SBIR/STTR) or credit-based programs (e.g., SBIC), fostering a robust VC ecosystem can unlock entrepreneurial potential among financially constrained but high-potential innovators.

Furthermore, the findings imply that VC selectively invests in regions and industries that benefited from public R&D, thus promoting localized entrepreneurial community. By aligning public support for basic and applied research with VC-driven commercialization, policymakers can bridge the gap between scientific discoveries and commercial deployment.

[17] https://sifted.eu/articles/pension-reforms-uk-investment-news

References

Airoldi, Anna, and Petra Moser, 2024, Inequality in Science: Who Becomes a Star?, *NBER Working Paper* .

Akcigit, Ufuk, and William R. Kerr, 2018, Growth through heterogeneous innovations, *Journal of Political Economy* 126.

Antolin-Diaz, Juan, and Paolo Surico, 2022, The Long-Run Effects of Government Spending, *CEPR Discussion Papers* .

Ash, Elliott, and Stephen Hansen, 2023, Text Algorithms in Economics.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat, 2019, Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules, *Review of Economic Studies* 86.

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

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

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.

Black, Sandra E., and Philip E. Strahan, 2002, Entrepreneurship and Bank Credit Availability, *Journal of Finance* 57, 2807–2833.

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.

Christensen, Clayton M., 2011, *The Innovator's Dilemma*.

Cornaggia, Jess, Yifei Mao, Xuan Tian, and Brian Wolfe, 2015, Does banking competition affect innovation?, *Journal of Financial Economics* 115, 189–209.

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

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* .

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf, 2018, Cost of experimentation and the evolution of venture capital, *Journal of Financial Economics* 128.

Florida, Richard L., and Martin Kenney, 1988, Venture capital-financed innovation and technological change in the USA, *Research Policy* 17.

Goldman, Jim, Rajkamal Iyer, and Ramana Nanda, 2022, Amplifying the Fiscal Multiplier: The Role of Banks, *SSRN Electronic Journal* .

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.

Gompers, Paul a, 1994, The Rise and Fall of Venture Capital, *Business and Economic History* 23.

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

Gompers, Paul A., Josh Lerner, Margaret M. Blair, and Thomas Hellmann, 1998, What Drives Venture Capital Fundraising?, *Brookings Papers on Economic Activity. Microeconomics* 1998.

Greenspan, Alan, and Adrian Wooldridge, 2018, *Capitalism in America: A history*.

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

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.

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

Kantor, Shawn, and Alexander T Whalley, 2024, Moonshot: Public R&D and Growth, *SSRN Electronic Journal* .

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.

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

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.

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 2023 10:1* 10, 1–22.

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

Moretti, Enrico, Claudia Steinwender, and John Van Reenen, 2023, The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers, *Review of Economics and Statistics* .

Nanda, Ramana, and Matthew Rhodes-Kropf, 2017, Financing risk and innovation, *Management Science* 63.

Nanda, Ramana, and Jesper B. Sørensen, 2010, Workplace peers and entrepreneurship, *Management Science* 56.

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

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.

Rezaei, Roham, and Yufeng Yao, 2024, Venture Capital Response to Government-Funded Basic Science, *SSRN Electronic Journal* .

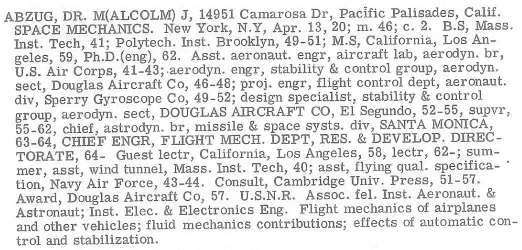
Robb, Alicia M., and David T. Robinson, 2014, The Capital Structure Decisions of New Firms, *Review of Financial Studies* 27, 153–179.

Samila, Sampsa, and Olav Sorenson, 2011, Venture capital, entrepreneurship, and economic growth, *Review of Economics and Statistics* 93.

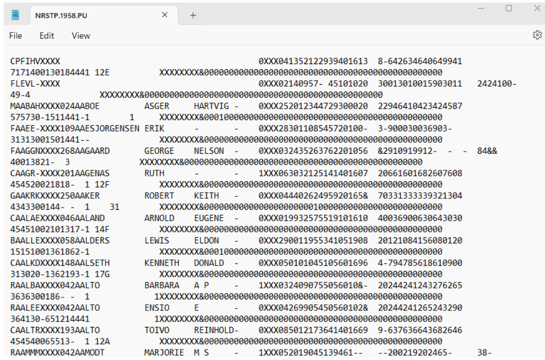
Figures

Figure 1: Examples of the Raw Data

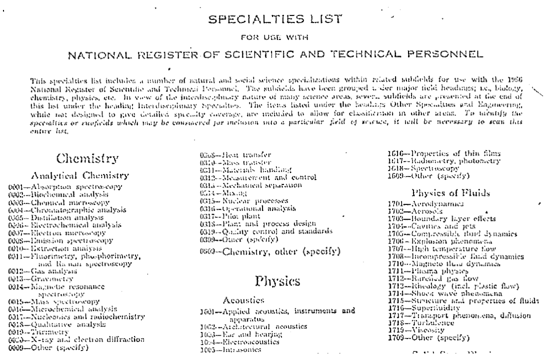
((**a**)) AWS



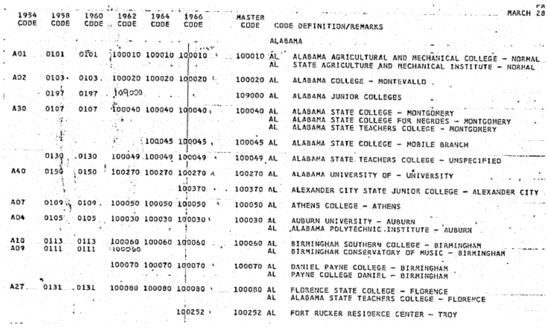
((**b**)) NRSTP: Entries



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

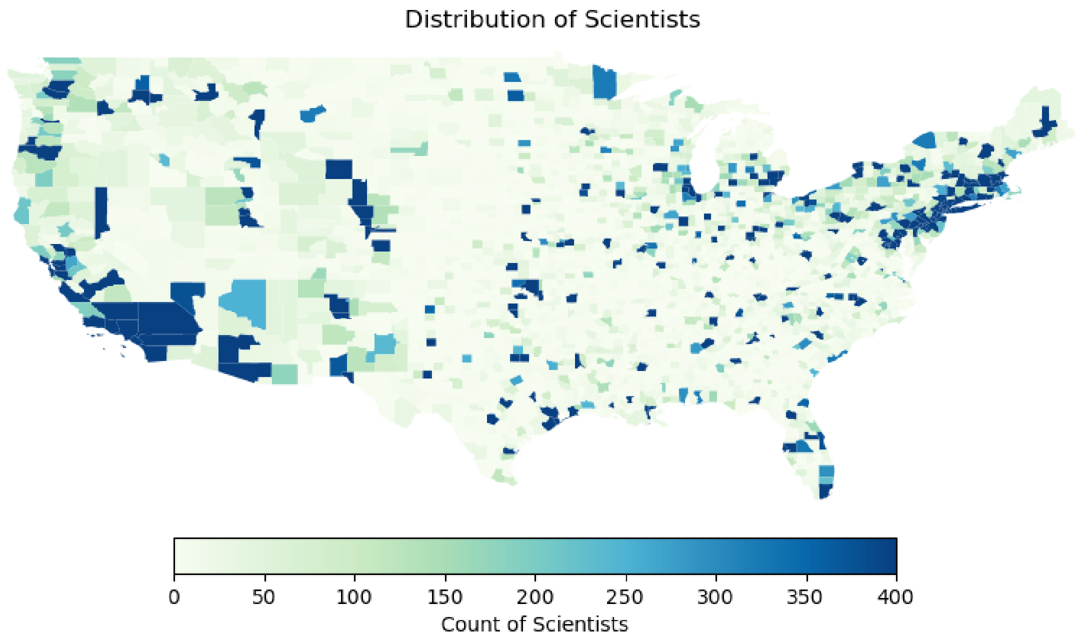


((**d**)) NRSTP: University Codebook



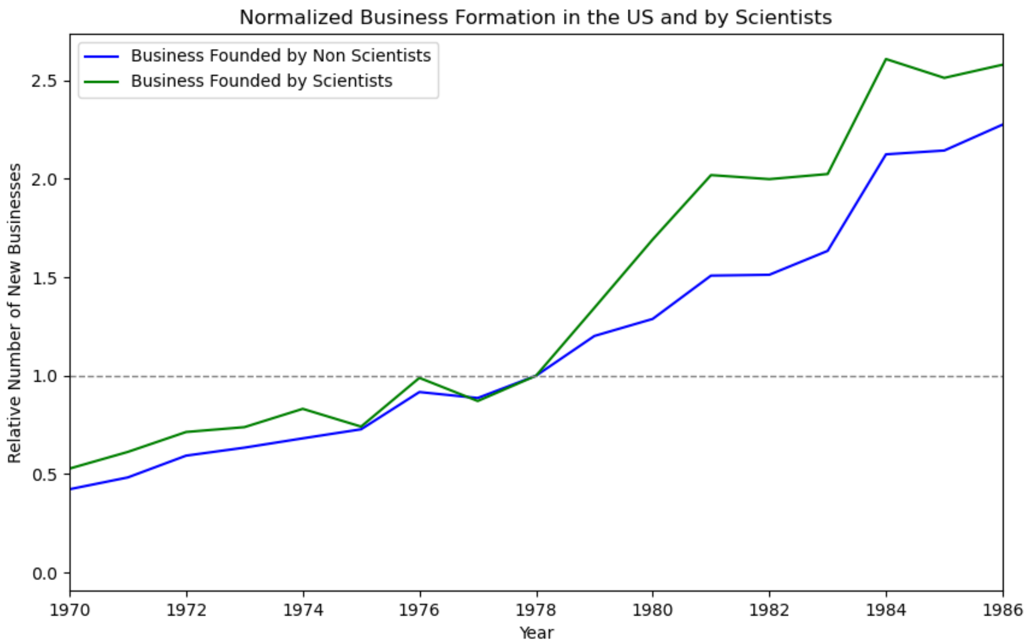
Notes: (a) shows the example of an AMS entry. Dr. Malcolm J. Abzug (April 13, 1920), an expert in space and flight mechanics, held prominent roles in aerodynamics, missile systems, and space research. Educated at MIT (Bachelor), Polytechnic Institute of Brooklyn (Master), and UCLA (PhD), he contributed significantly to Douglas Aircraft Co. and US Air Corps. His research focused on flight mechanics, fluid mechanics, and control systems. (b) shows the raw dataset from the NRSTP. Each line represents an entry of scientists. 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



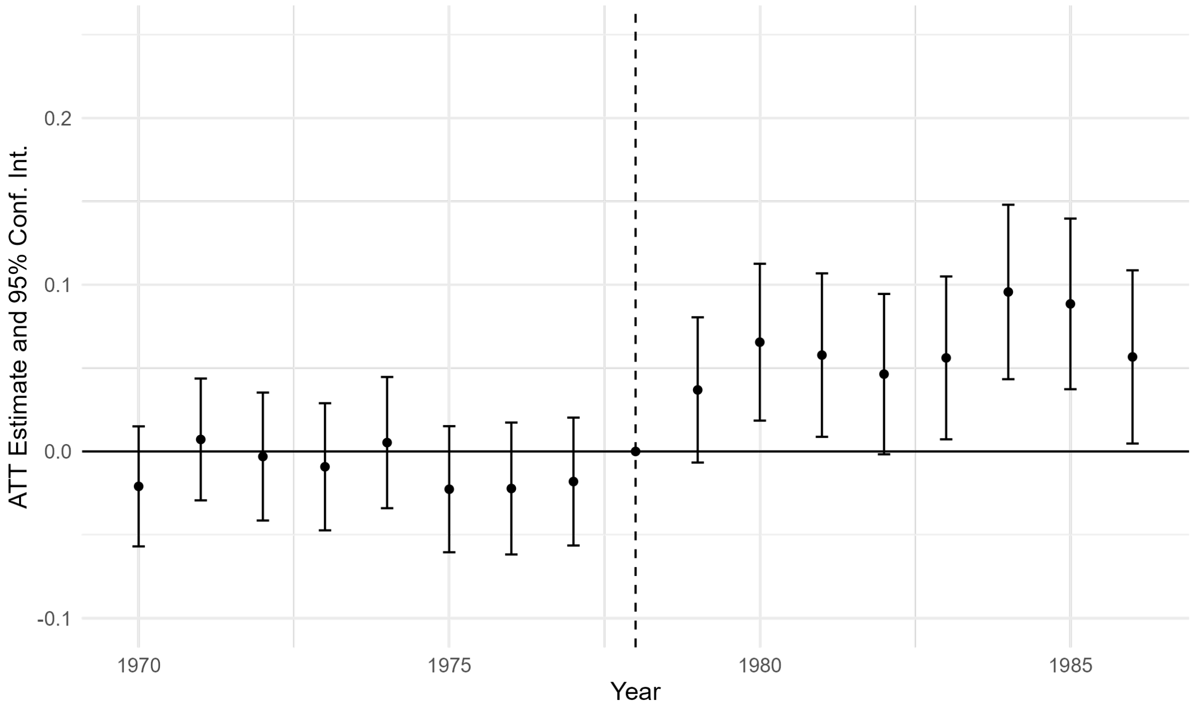
Notes: This figure plots the geographical distribution of scientists, using county delineations from the 1990 Census. The historical county FIPS crosswalk follows Eckert et al. (2020). The scientist counts are weighted to account for differences in population weights between 1990 and 2010. For visualization purposes, the color scale is capped at 500. Counties with more than 500 scientists are represented using the same color as those with exactly 500 scientists.

Figure 3: Business Formation Trend in the US



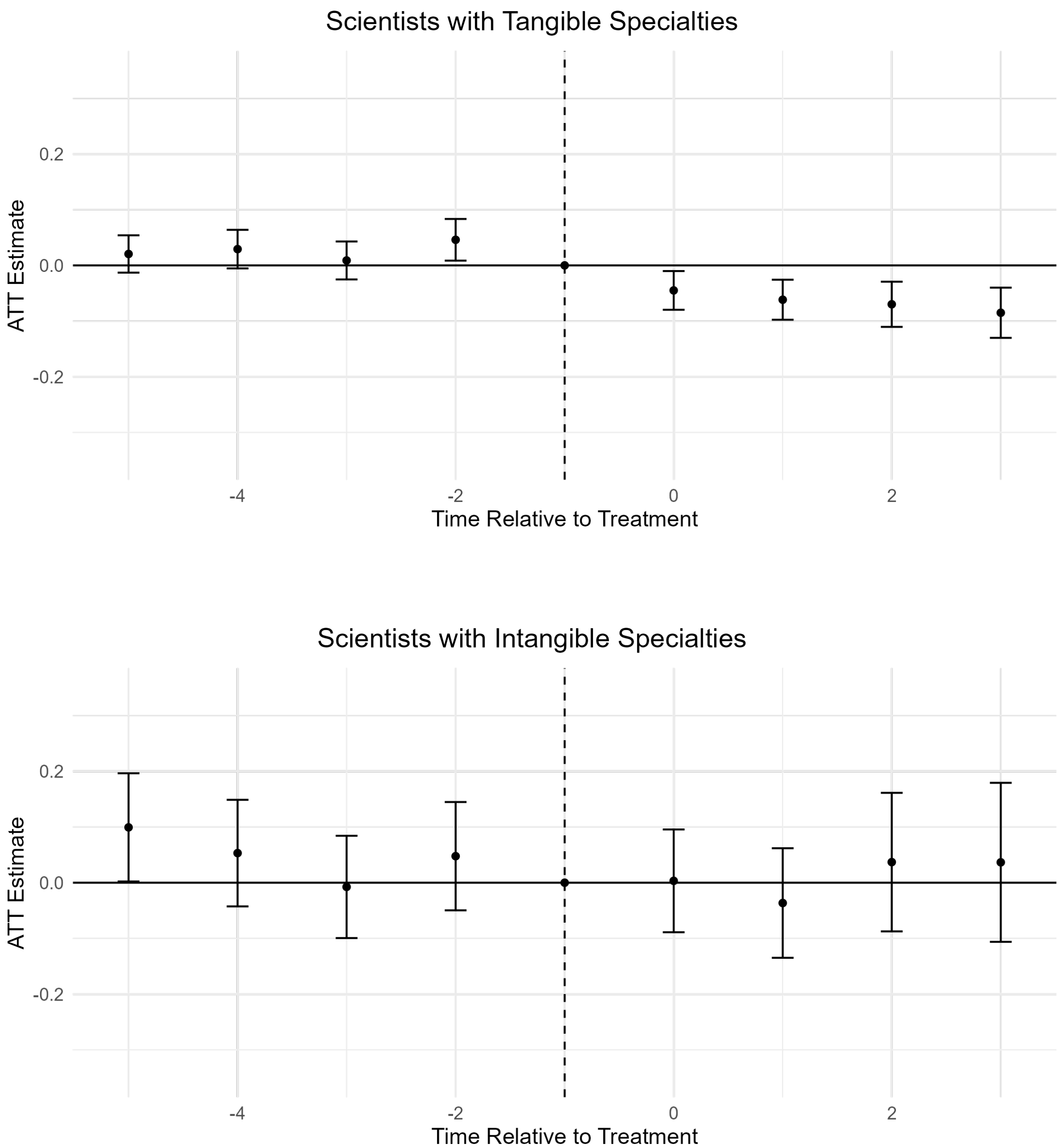
Notes: This figure plots the number of businesses incorporated in the US and those founded by scientists. Data is from OpenCorporates. Business formation counts are normalized to 1978 (set to 1) for comparison. The total US business formation includes all newly incorporated businesses, while scientist-founded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

Figure 4: Business Formation Difference between Intangible and Tangible Scientists



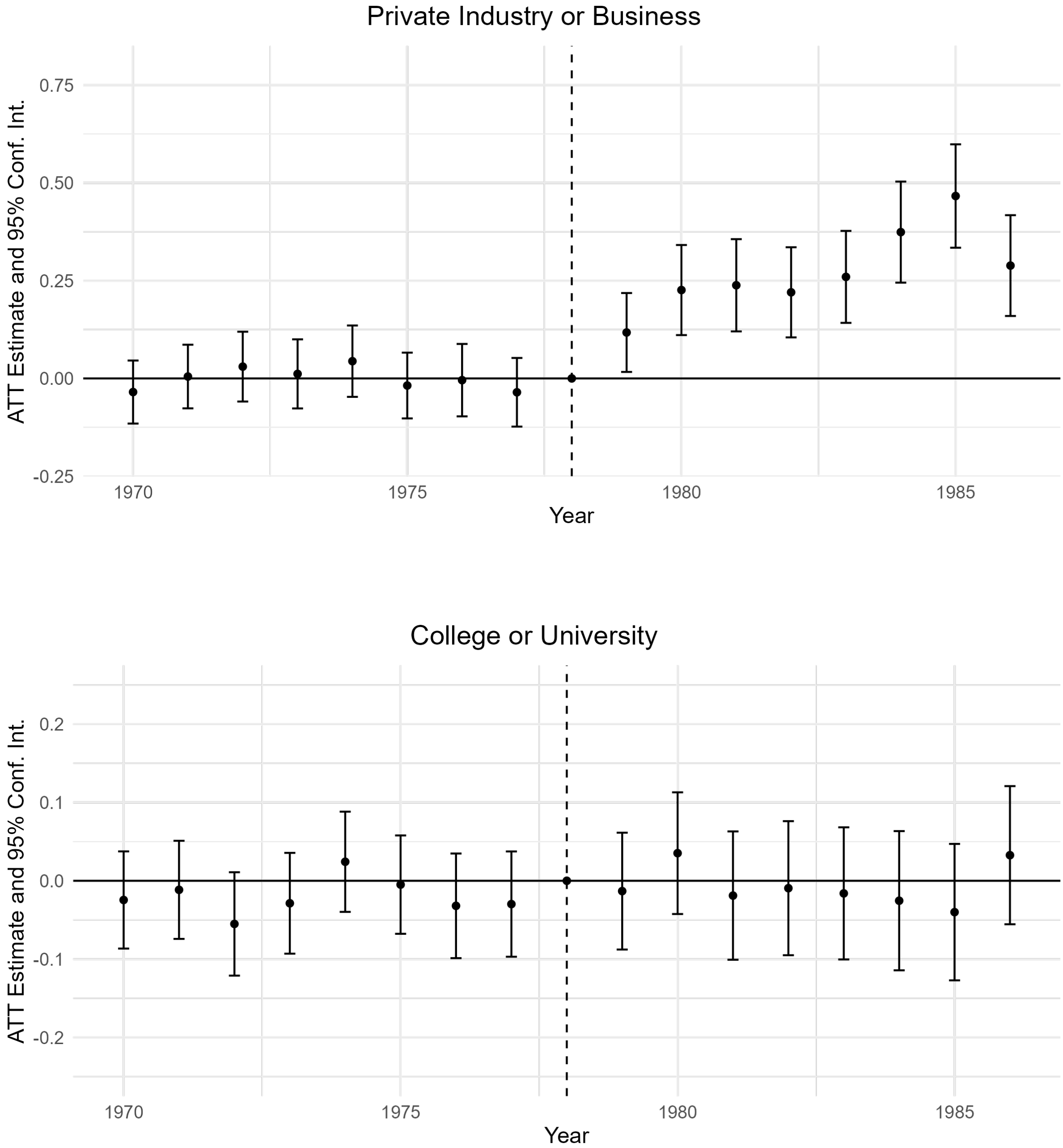
Notes: This figure displays the coefficients from the difference-in-differences estimation from column (4) in the Table 3. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 5: Event Study of Intrastate Branching Deregulation on Business Formation



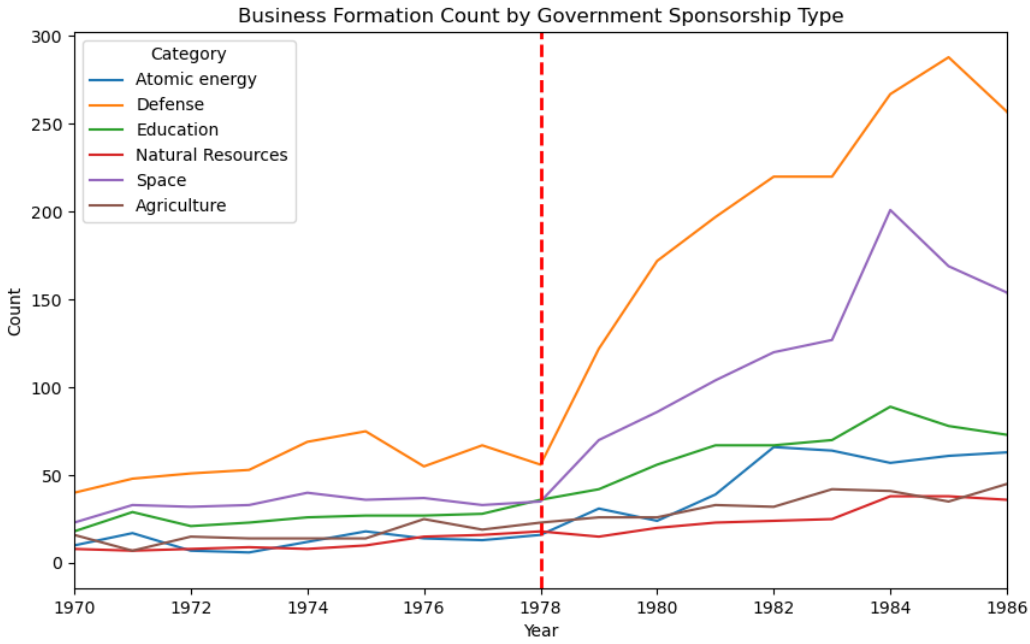
Notes: This figure plots the estimated coefficients from the event study based on Equation (2), with the sample stratified by scientists' work specialties. It highlights the heterogeneous effects of intrastate bank deregulation on business formation across different specialties. The vertical lines denote 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 with tangible specialty is -0.128, with a standard error of 0.0166.

Figure 6: Business Formation of Other Employer Type



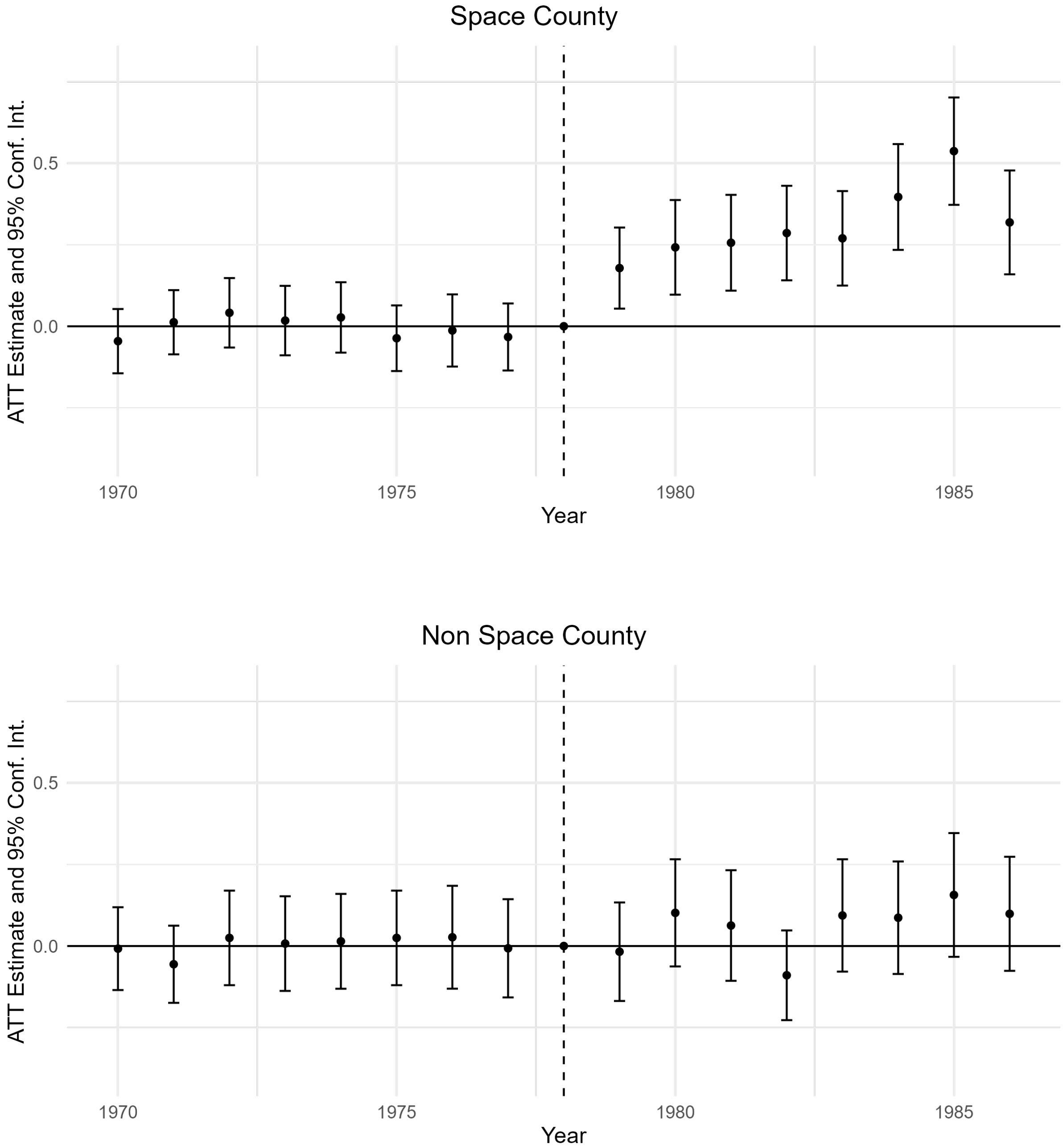
Notes: This figure displays the coefficients from the difference-in-differences estimation of Columns (1) and (2) in Table 5. It illustrates the heterogeneous treatment effects based on the type of employer for scientists and engineers. The vertical lines represent the 95% confidence intervals for the coefficient estimates.

Figure 7: Business Formation Count by Government Sponsorship Type



Notes: This figure plots the business formation of scientists and engineers during 1970-1986. The scientists are classified based on their government-sponsored program, as self-reported in the NRSTP.

Figure 8: Business Formation Difference between Space and Non-Space Scientists



Notes: This figure displays the coefficients from the difference-in-differences estimation of Columns (2) and (4) in Table 8. It shows that industry scientists living in the Space Counties are more responsive to the increasing supply of VC. 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	25%	Mean	50%	75%	Max	Std. Dev.
BizCount	475,334	0.00	0.00	0.0591	0.00	0.00	57.00	0.5517
StartBusiness	475,334	0.00	0.00	0.0315	0.00	0.00	1.00	0.1746
PatCount	475,334	0.00	0.00	0.3879	0.00	0.00	356.00	2.7892
FilePatent	475,334	0.00	0.00	0.0904	0.00	0.00	1.00	0.2868
PublicationCount	475,334	0.00	0.00	2.1253	0.00	0.00	1,273.00	15.1876
PubPaper	475,334	0.00	0.00	0.0947	0.00	0.00	1.00	0.2928

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 has started at least one firm. PatCount is the number of patents where the scientist is listed as an inventor. FilePatent equals one if a scientist has filed at least one patent. PublicationCount is the number of journal publications authored by the scientist. PubPaper equals one if a scientist has published at least one journal article.

Table 2: Top Intangible and Tangible Specialties

Intangible Specialties	Tangible Specialties
Information Retrieval	Astronautical Engineering
Behavior	Textile Engineering
Geography	Aeronautical Engineering
Monetary and Fiscal Theory	Marine Engineering
Communication Science	Chemical Engineering
Operations Research	Aerospace Engineering
Ecology	Hydraulic and Sanitary Engineering
Taxonomy	Petroleum Engineering
Epidemiology	Materials Engineering
Information Science	Electrochemical Engineering
Communication	Ceramic Engineering
Information System Design	Civil Engineering
Game Management	Mechanical Engineering
Theory and Practice of Computation	Civil and Structural Engineering
Operations Analysis	Metallurgical Engineering
Evolution	Plastics Engineering
Genetics and Animal Behavior	Mechanical and Industrial Engineering
Programmed Learning	Metallurgy and Materials Engineering
Insect Ecology	Material Engineering

Notes: The table reports the top intangible and tangible specialties based on textual similarity. Astronautical Engineering Engineering has the highest difference between the tangible and intangible scores, indicating that it is highly tangible. In contrast, Information Retrieval has the lowest difference between the tangible and intangible scores, suggesting it is the most intangible specialty.

Table 3: ERISA and the Business Formation of Scientists

Dependent Variable:	**StartBusiness**			
	(1)	(2)	(3)	(4)
Constant	0.1018***	-0.0404***		
	(0.0033)	(0.0051)		
Post1979	0.1279***	0.1241***		
	(0.0057)	(0.0057)		
Intangible	0.0283***	-0.0110*	-0.0109*	
	(0.0056)	(0.0056)	(0.0056)	
Post1979 × Intangible	0.0786***	0.0748***	0.0747***	0.0469***
	(0.0100)	(0.0100)	(0.0100)	(0.0099)
Controls		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	4,206,548	4,177,470	4,177,470	4,177,470
R^2	0.00042	0.00382	0.00388	0.13240

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 LLM classification. Post 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: ERISA and the Innovation Activity of Scientists

Dependent Variable:	FilePatent		PubPaper	
	(1)	(2)	(3)	(4)
Intangible	-0.3413***		-0.0320	
	(0.0127)		(0.0342)	
Post1979 × Intangible	0.1616***	0.1585***	-0.0514**	-0.0492**
	(0.0102)	(0.0101)	(0.0245)	(0.0246)
Controls	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Individual FE		Yes		Yes
Observations	2,706,894	2,706,894	4,177,470	4,177,470
R^2	0.00118	0.39565	0.00100	0.41246

Notes: This table reports the difference-in-differences estimates of the ERISA effect on patenting and publishing activities by scientists. The estimate for patenting activity is from 1976 to 1986. Due to the USPTO's publication of patent inventor data beginning in 1976 and inconsistencies in historical inventor data around that year, I excluded panel data before 1976. The dependent variable *FilePatent* is a binary indicator of whether a scientist filed a patent in a given year. The dependent variable *PubPaper* is a binary indicator of whether a scientist published a journal article in a given year. The estimate for publishing activity is from 1970 to 1986. *Intangible* is a binary variable indicating whether the scientist's work specialty is classified as intangible based on LLM classification. Post 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: Mechanism Test: Type of Employers

Dependent Variable:	StartBusiness			
	(1) Private Industry	(2) College and University	(3) Federal Government	(4) Self Employed
Post1979 × Intangible	0.1564***	0.0075	0.0275	0.0527
	(0.0236)	(0.0160)	(0.0416)	(0.0955)
Controls	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,577,200	1,245,118	183,766	78,078
R^2	0.13808	0.12890	0.11920	0.11963

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 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 LLM classification. Post 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: Mechanism Test: Employees Productivity and Spinouts

Dependent Variable:	**StartBusiness**			
	Panel A: Scientists Working in Private Industry or Sector			
	(1) No Patent	(2) Have Patent	(3) No Publication	(4) Have Publication
Post1979 × Intangible	0.1293***	1.275***	0.1457***	0.4221**
	(0.0229)	(0.2240)	(0.0236)	(0.1954)
Controls	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,254,992	322,208	1,546,915	30,285
R^2	0.12663	0.17352	0.13641	0.17090
	Panel B: Scientists Working in College or University			
	(5) No Patent	(6) Have Patent	(7) No Publication	(8) Have Publication
Post1979 × Intangible	-0.0084	1.732***	-0.0007	0.0446
	(0.0153)	(0.3954)	(0.0170)	(0.0441)
Controls	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	1,217,849	27,269	1,018,650	226,468
R^2	0.11938	0.19682	0.12759	0.13274

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 have filed a patent/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 LLM classification. Post 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: Mechanism Test: Flow of VC Investment

Dependent Variables	**Log(DealSize)**		**DealCount**	
	(1)	(2)	(3)	(4)
SpaceCountyIndustry	-0.0107**	-0.0053**	-0.0023**	-0.0009***
	(0.0053)	(0.0021)	(0.0011)	(0.0003)
Post1979 × SpaceCountyIndustry	0.0400***	0.0314***	0.0102*	0.0079***
	(0.0144)	(0.0059)	(0.0053)	(0.0017)
County FE	Yes		Yes	
Year FE	Yes		Yes	
Industry FE	Yes	Yes	Yes	Yes
County-Year FE		Yes		Yes
Observations	302,157	302,157	302,157	302,157
R^2	0.08018	0.14392	0.09204	0.15103

Notes: This table reports the difference-in-differences estimates of the ERISA effect on the venture capital flow. The analysis is restricted to early-stage deals with an investment stage categorized as Seed, Early Stage, or VC Partnership. The variable for deal size represents the natural logarithm of the disclosed equity contribution (in USD). *SpaceCountyIndustry* is an indicator variable reflecting a county-industry's being above median in terms of the similarity between the technologies present in pre-1958 patents and the National Intelligence Estimates of Soviet Space Capabilities between 1958 and 1992 (the Space Capability Score), as described on Kantor and Whalley (2024). The results remain consistent whether or not a logarithmic transformation is applied. Deal count refers to the total number of deals within each county. Standard errors are clustered by county. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8: Mechanism Test: Space Counties and Residence Locations

Dependent Variable:	**StartBusiness**			
	(1)	(2)	(3)	(4)
	Space County		Non Space County	
	Full Sample	Private Industry	Full Sample	Private Industry
Post1979 × Intangible	0.0637***	0.1711***	-0.0085	0.0493
	(0.0138)	(0.0315)	(0.0133)	(0.0316)
Controls	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	2,585,601	944,505	1,408,307	587,713
R^2	0.14143	0.14864	0.10006	0.10284

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 county where the scientists reside and their type of employer. Space County is defined based on the space score developed by (Kantor and Whalley, 2024), which proxies the likelihood of the county receiving investment during the Space Race. 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 LLM classification. Post equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$

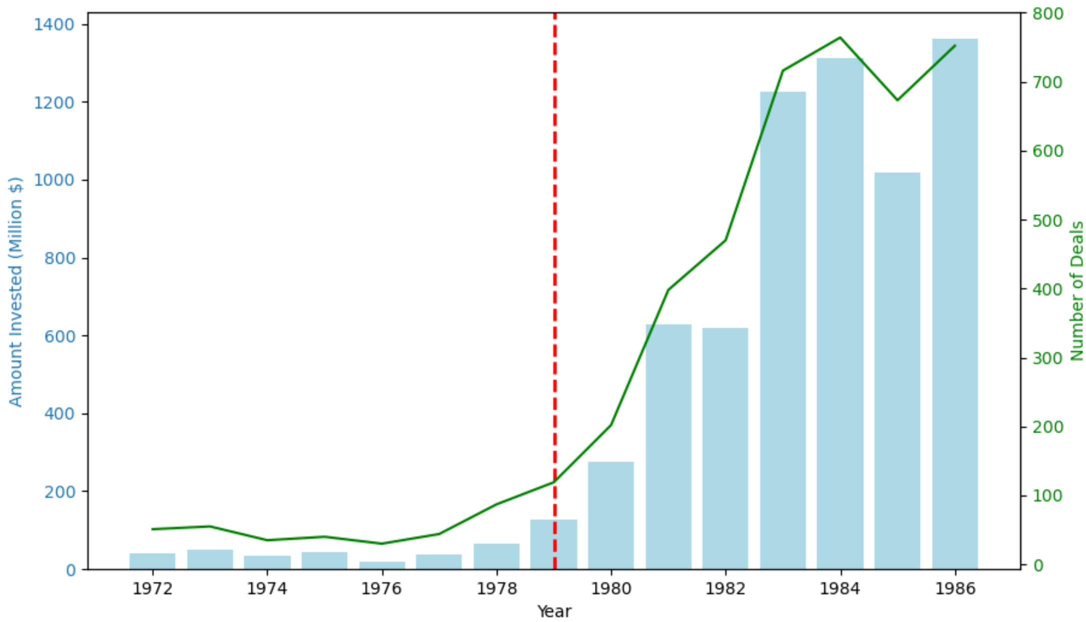
A Appendix Figures

Figure A1: Original Document of 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.

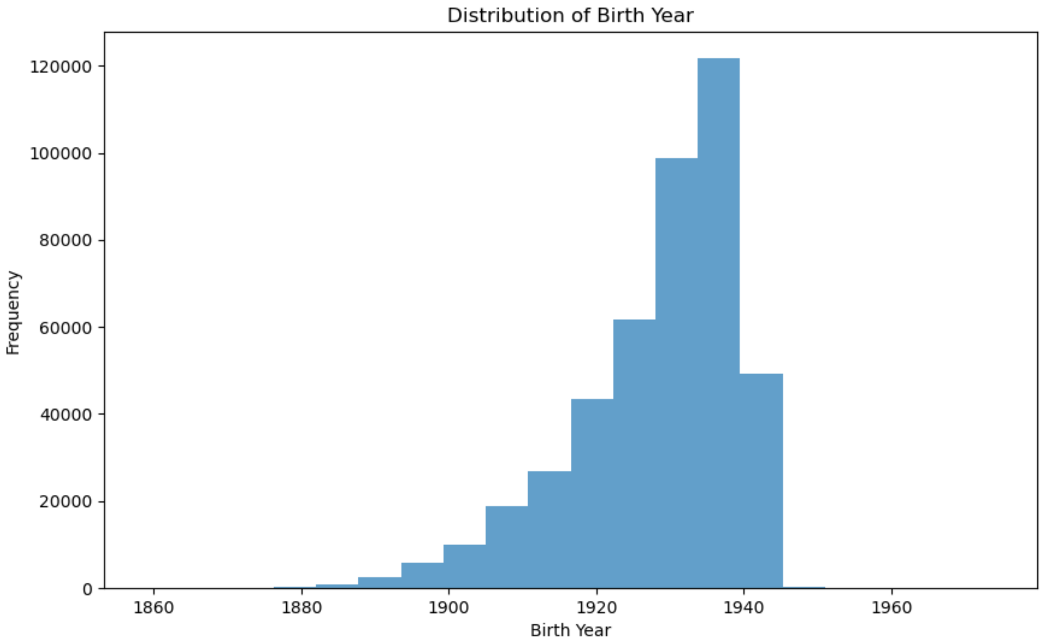
Notes: This graph shows the 29 Code of Federal Regulations Part 2550, 1979. This is the final regulation on the *Rules and Regulations for Fiduciary Responsibility; Investment of Plan Assets Under the "Prudence" Rule*. The amendment was published in the Federal Register on June 26, 1979. Federal agencies typically begin drafting amendments well before 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



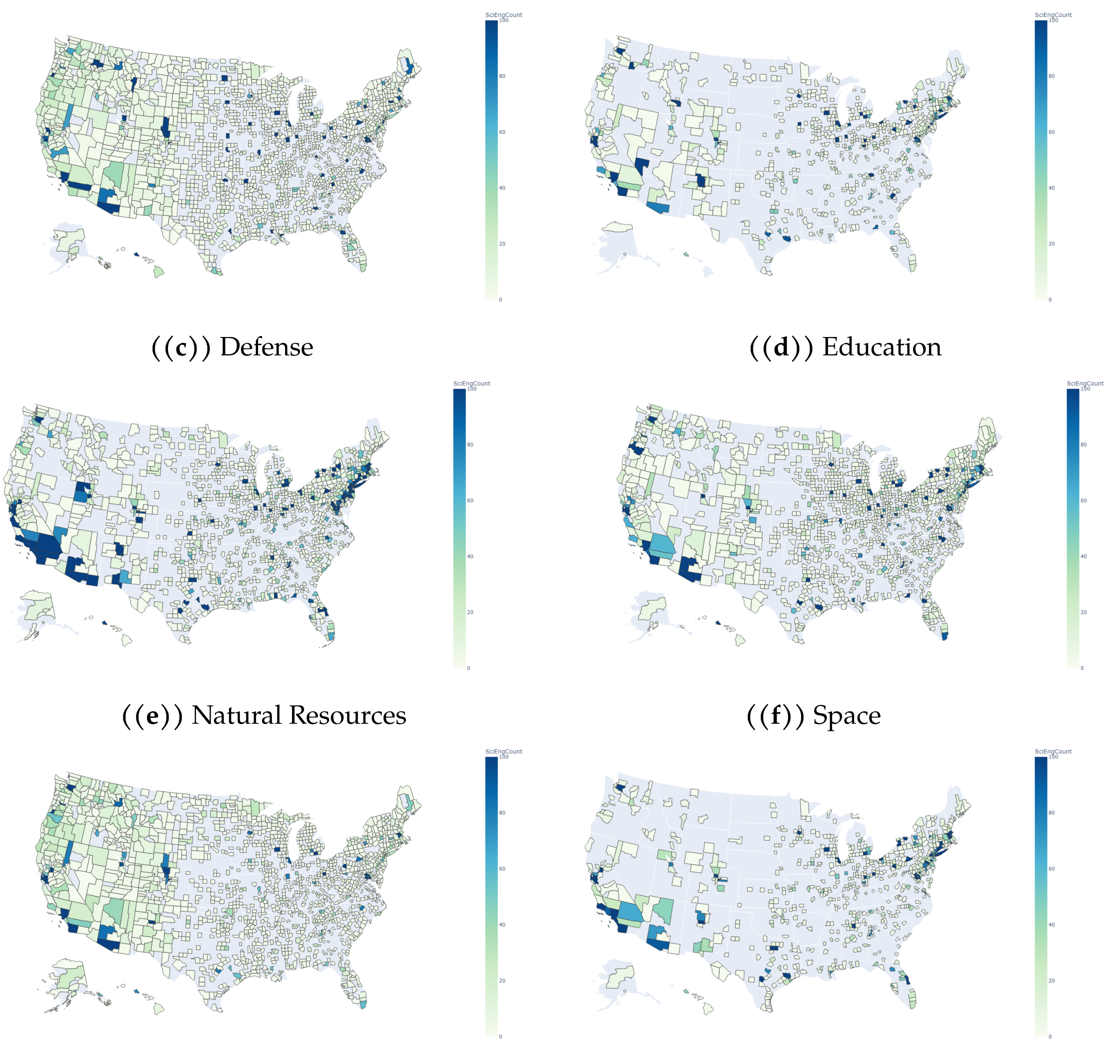
Notes: This figure plots the total amount of VC investment and the number of deals in the US. Note that these values are underestimated due to incomplete data coverage in the dataset, particularly for the 1980s. Also, many of the deals did not disclose the deal sizes. Data comes from Venture Economics, a repository of information widely recognized in the field of economics, particularly focusing on 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 et al., 2018).

Figure A3: Birth Year of the Scientists



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

Figure A4: Geographical location of government-funded scientists and engineers



((**a**)) Agriculture

((**b**)) Atomic Energy

((**c**)) Defense

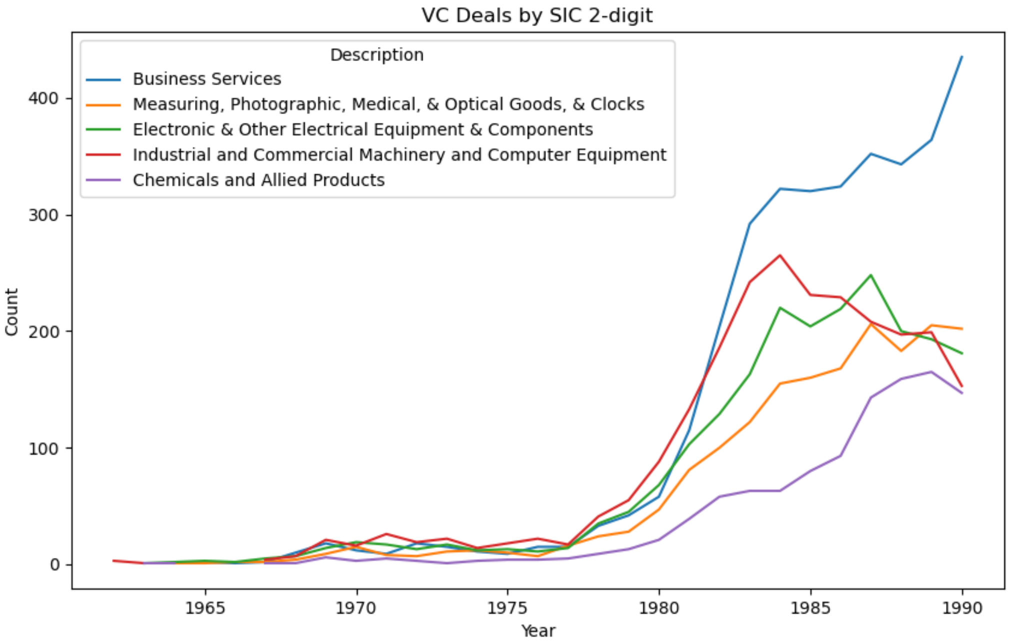
((**d**)) Education

((**e**)) Natural Resources

((**f**)) Space

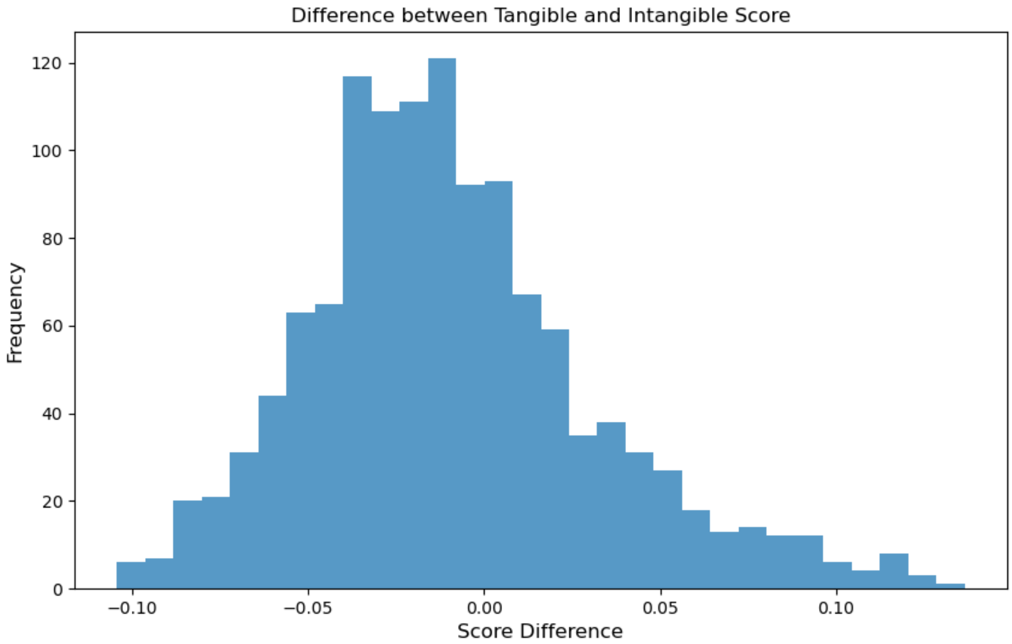
Notes: The graph presents the geographical distribution of scientists and engineers as recorded in the 1962, 1964, 1966, and 1968 NRSTP. These surveys from the specified years include data on whether a respondent is involved in government-funded projects in specific areas. The visualization uses color to represent the count of scientists in each county, with the color intensity indicating the number of scientists present. To standardize the comparison, the color scale is capped at a maximum count of 100. For counties where no respondents are located, their boundaries are not outlined, distinguishing them from those with recorded scientists and engineers.

Figure A5: Eary-Stage VC Deals by Industries



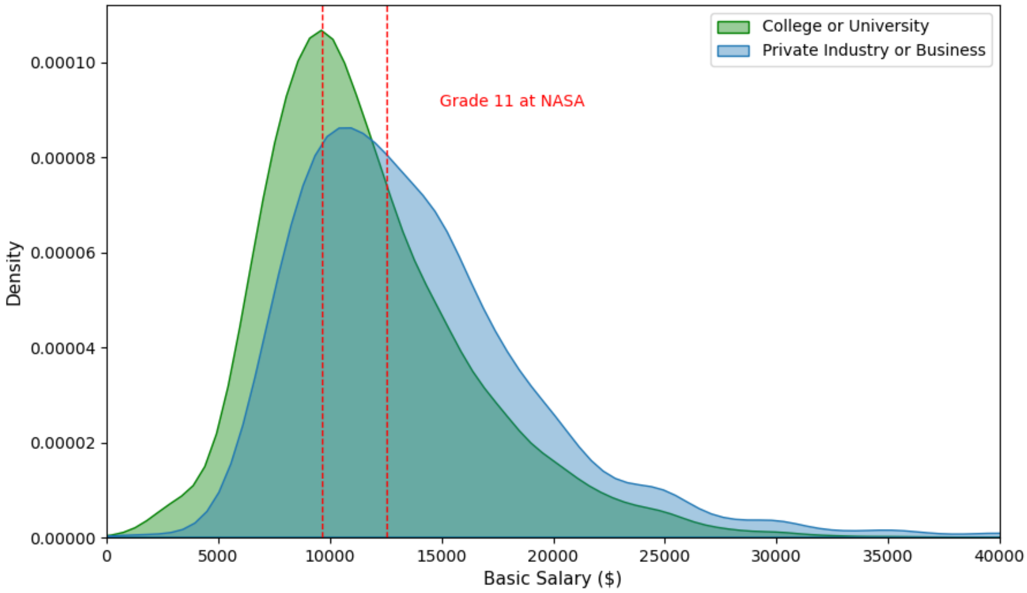
Notes: This figure plots the number of VC deals from 1960 to 1990 based on 2-digit SIC codes. The top five industries by deal count in 1990 are selected. Data comes from Venture Economics.

Figure A6: Difference between Tangible and Intangible Scores



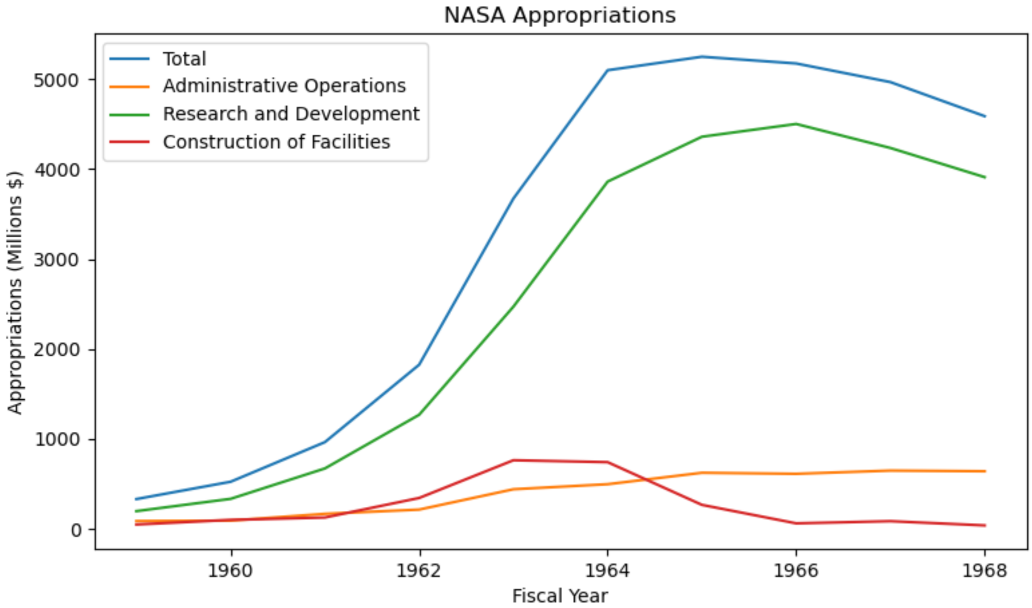
Notes: This figure plots the distribution of differences between tangible and intangible similarity scores for each specialty. The works specialty of scientists in the sample are more intangible in general. This data is plotted at the specialty level, not the individual level.

Figure A7: Basic Salary and NASA General Schedule Grade for Employees



Notes: This graph presents the distribution of scientists' annual base salaries. It includes a reference line derived from NASA's General Schedule Salary Rates in 1968, drawing from the NASA Historical Data Book. There are a total of 18 grades, with Grade 18 representing the highest salary rate. According to the Historical Data Book, GS-14 salaries span from \$815,841 to \$820,593, GS-13 from \$813,507 to \$817,557, GS-12 from \$811,461 to \$814,899, GS-11 from \$809,657 to \$812,555, and GS-10 from \$808,821 to \$811,467. For illustration purposes, salaries above 40,000 are dropped from the graph but are included in the density analysis.

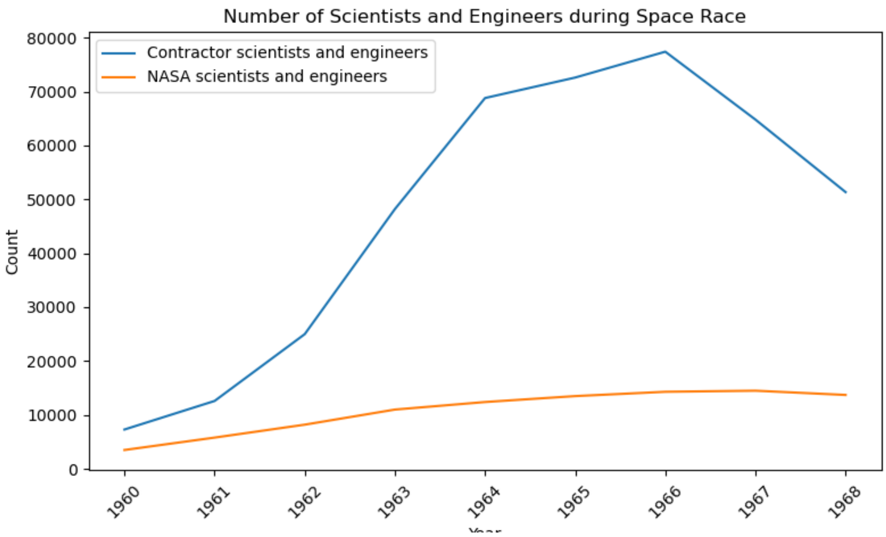
Figure A8: NASA Appropriations 1959-1968



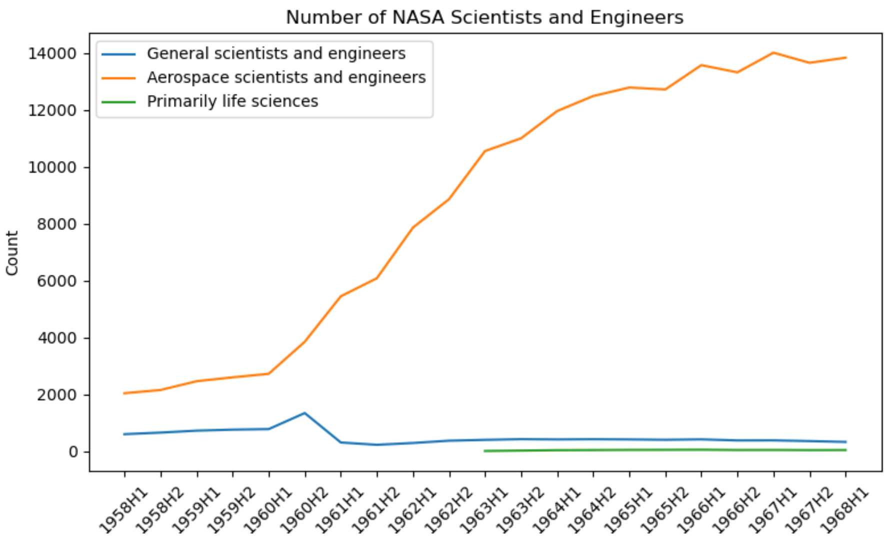
Notes: Data are drawn from NASA Historical Data Books page 116. The data for FY 1968 is as of June 30. During its first decade, NASA spent (obligated) just over $32 billion. This sum represented over three percent of the money spent by the federal government.

Figure A9: NASA Technical Workforce during the Space Race

((**a**)) NASA in house and contractors



((**b**)) NASA in house



Notes: Data are drawn from NASA Historical Data Books. General Scientist and Engineers include professional positions in physical sciences, engineering, and mathematics that are not specifically associated with aerospace technology. Aerospace Scientific and Engineering professional scientific and engineering positions requiring Aero-Space Technology (AST) qualifications. This category encompasses professional roles engaged in aerospace research, development, operations, and related work, including the development and operation of specialized facilities and supporting equipment. Life Science includes professional life science positions that do not require AST qualifications. This category includes medical officers and other positions performing professional work in psychology, the biological sciences, and professions that support the science of medicine, such as nursing and medical technology.

B Appendix Tables

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

Level of Highest Degree	Count
Bachelor	146,490
Master	133,482
MD	11,026
PhD	169,262
PhD+	2,342

Notes: Compared to the AMS, NRSTP offers a broader view of the workforce. 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. The 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 highest degree of scientist and engineers. 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 income distribution of the scientist and compare it with the US general population. The PSZ data is from the 2022 version of TB3 from Distributional National Accounts by Piketty et al. (2018) https://gabriel-zucman.eu/usdina/

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 scientists and engineers. Over the years, the classification of employer types has become increasingly granular. I manually created a crosswalk file to harmonize these classifications. In 1970, the category "State, local, or other government (except educational institution)" includes entities such as the USPHS Commissioned Corps, U.S. Weather Bureau, State Government, International Agencies, and Other Government Agencies. Research centers managed by 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. 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 specialty of the scientists and engineers. The data comes from both NRSTP and AMS. The NRSTP part originates from the "Professional Characteristics section" of the questionnaire, where respondents were asked to identify the specialties in which they believe they have 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 part comes from the list of academic disciplines provided by the AMS.

Table A7: Government Sponsored Work Specialties

Panel A: Defense Programs	
Work Specialty	Share of Scientists
Engineering Psychology	0.68
Aeronautical Engineering	0.60
Human Engineering	0.60
Meteorological Instrumentation	0.56
Acoustics	0.55
Network Engineering	0.53
Electricity and Magnetism	0.52
Aeronautical and Astronautical Engineering	0.52
Synoptic Meteorology	0.51
Geodesy	0.51
Panel B: Space Programs	
Work Specialty	Share of Scientists
Electronics Engineering	0.75
Environmental Engineering	0.71
Solar/Planetary Specialties	0.69
Engineering of General	0.53
Material Engineering	0.52
Aeronautical and Astronautical Engineering	0.51
Aeronautical Engineering	0.51
Energy Conservation Programs	0.51
Engineering Science	0.48
Astronomy	0.43

Notes: This table presents the top scientific specialties associated with government-sponsored programs. The data is sourced from the NRSTP, where scientists self-report their participation in government funding programs. The reported share represents the proportion of scientists within each specialty who receive support from a specific program. Panel A lists the leading specialties within government defense programs, while Panel B highlights those most associated with government space funding.

Table A8: Correlation Matrix

	BizCount	FirmPatCount	PatCount	PaperCount	CitationCount	TeamSize	InstitutionCount	PatentCount
BizCount	1.000							
FirmPatCount	0.058	1.000						
PatCount	0.017	0.002	1.000					
PaperCount	0.005	0.003	0.002	1.000				
CitationCount	0.007	0.002	0.003	0.035	1.000			
TeamSize	0.019	0.006	0.025	0.167	0.062	1.000		
InstitutionCount	0.009	0.017	-0.008	0.028	0.028	0.372	1.000	
PatentCount	-0.004	-0.000	0.034	0.220	0.072	0.071	0.011	1.000

Notes: This table presents the correlation matrix between key variables. All variables are at the individual level. BizCount is the number of businesses formed by the scientist. FirmPatCount is the number of patents owned by the firm founded by the scientist. PatCount is the number of patents where the scientist is listed as an inventor. PaperCount is the number of journal publications authored by the scientist. CitationCount is the average number of citations received by the scientist's publications. TeamSize is the average number of researchers per paper. InstitutionCount is the number of institutions listed in the paper. PatentCount is the number of citations received from patents registered with the USPTO and EPO.

Table A9: Dictionary of Tangible and Intangible Specialties

Tangible Specialties	Intangible Specialties
Materials Science	Computer Science
Metallurgy	Information Theory
Mechanical Engineering	Artificial Intelligence
Chemical Engineering	Software Engineering
Physics	Cognitive Science
Industrial Chemistry	Telecommunications
Polymer Science	Cybernetics
Ceramics Engineering	Linguistics
Electrical Engineering	Economics
Civil Engineering	Intellectual Property Law
Geology	Information Systems
Mineralogy	Operations Research
Mining Engineering	Management Science
Petroleum Engineering	Psychology
Manufacturing Engineering	Knowledge Management
Aeronautical Engineering	Human-Computer Interaction
Textile Engineering	Sociology
Construction Engineering	Educational Technology
Automotive Engineering	Decision Theory
Marine Engineering	Business Administration

Notes: The table reports the dictionaries created by the GPT-4o. The prompt I used was "give me 20 scientist specialties that are related to tangible assets in 1970s and 1980s. only give me the words as a python list format, no explanation." The words are stored in two separate lists, the tangible and intangible dictionaries, respectively. These lists are embedded into two vectors using SciBERT. The similarity between a scientist's work specialty and these vectors is then calculated to determine its classification.

Table A10: 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 specialty intangible classification of the same individuals who appear in both the 1962 NRSTP survey and the 1968 survey. The values represent the counts of observations transitioning between categories. Individuals whose work specialty changed from 1 to "not able to define" were dropped.

Table A11: Differences between Tangible and Intangible Scientits

Variable	Tangible	Intangible	Diff in Mean	t-statistic
Female	0.06	0.15	-0.08 ***	-72.76
Year of Highest Degree	1953.50	1955.77	-2.26 ***	-57.86
Basic Salary	13,008.04	12,932.29	75.75 ***	3.05
Gross Income	13,249.67	13,656.54	-406.87 ***	-13.29
Govt. Agriculture	0.09	0.05	0.04 ***	35.42
Govt. Atomic Energy	0.06	0.04	0.01 ***	11.80
Govt. Defense	0.12	0.19	-0.07 ***	-42.09
Govt. Education	0.06	0.16	-0.10 ***	-71.43
Govt. Natural Resources	0.05	0.02	0.03 ***	41.50
Govt. Space	0.05	0.08	-0.03 ***	-22.24
EmployerFirm	0.54	0.24	0.30 ***	171.27
EmployerGov	0.04	0.06	-0.02 ***	-27.48
EmployerMil	0.01	0.02	-0.00 ***	-9.05
EmployerUni	0.23	0.46	-0.23 ***	-131.66

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 Top Share of Intangible and Tangible Work Specialty Employees

Intangible Specialties	Tangible Specialties
Rohrer Hibler & Replogle	Dexter Corp
Informatics Inc	Devoe & Raynolds Co Inc
American Inst for Research	Fiberite Corp
Applied Data Research Inc	Richardson Co
Wyatt Co	Holland Suco Color Co
Morgan Guaranty Trust Co	Sheller Mfg Corp
Systems Dev Corp	Simoniz Co
Pacific Tech Analysts Inc	Crawford & Russell Inc
Philip Hankins & Co Inc	Blaw Knox
Computer Control Co	Catalytic Construction Co
Computing & Software Inc	Ash Stevens Inc
Harcourt Brace & World Inc	Jim Walter Research Corp
Scientific Data Systems	Sonoco Products Co
System Development Corp	Norda Essential Oil & Chemical Co
Computer Usage Co	Singmaster & Breyer
Austen Riggs Center	Reeves Brothers Inc
Touche Ross Bailey & Smart	Arthur G. McKee & Co
Pacific Mutual Life Insurance Co	Pratt & Lambert Inc
Menninger Foundation	Titanium Pigment Corp
American Inst for Res	Debell & Richardson 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: Robustness Check: Continuous Intangibility Scores

Dependent Variable:	**StartBusiness**			
	(1)	(2)	(3)	(4)
Constant	-0.0373	-0.0563		
	(0.0353)	(0.0357)		
Post1979	-0.2308***	-0.2124***		
	(0.0625)	(0.0627)		
IntangibleScore	0.1914***	0.0108	0.0110	
	(0.0449)	(0.0453)	(0.0453)	
Post1979 × IntangibleScore	0.5068***	0.4752***	0.4752***	0.2389***
	(0.0797)	(0.0799)	(0.0799)	(0.0793)
Controls		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	8,080,678	8,020,272	8,020,272	8,020,272
R^2	0.00039	0.00395	0.00401	0.13468

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. *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. Post 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 A14: Robustness Check: Continuous Tangibility Scores

Dependent Variable:		**StartBusiness**		
	(1)	(2)	(3)	(4)
Constant	0.1069***	-0.1313***		
	(0.0305)	(0.0311)		
Post1979	0.0964*	0.0979*		
	(0.0546)	(0.0547)		
TangibleScore	0.0087	0.1052***	0.1051***	
	(0.0384)	(0.0389)	(0.0389)	
Post1979 $\times$ TangibleScore	0.0918	0.0816	0.0817	0.0073
	(0.0690)	(0.0692)	(0.0692)	(0.0690)
Controls		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	3,247,153	3,224,835	3,224,835	3,224,835
R^2	0.00046	0.00379	0.00386	0.13234

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. Post 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 A15: Robustness Check: Intangible Specialty Definition with a Different Cutoff

Dependent Variable:	**StartBusiness**			
	(1)	(2)	(3)	(4)
Constant	0.1101***	-0.0403***		
	(0.0026)	(0.0039)		
Post1979	0.1615***	0.1552***		
	(0.0047)	(0.0046)		
Intangible_1	0.0077**	-0.0169***	-0.0169***	
	(0.0039)	(0.0039)	(0.0039)	
Post1979 $\times$ Intangible_1	0.0160**	0.0155**	0.0155**	0.0126*
	(0.0071)	(0.0071)	(0.0071)	(0.0071)
Controls		Yes	Yes	Yes
Year FE			Yes	Yes
Individual FE				Yes
Observations	8,080,678	8,020,272	8,020,272	8,020,272
R^2	0.00037	0.00394	0.00400	0.13468

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_1* is a binary variable indicating if the intangible score is higher than the tangible score based on the LLMs measures. Post equals one for years after 1978. All specifications include individual fixed effects and year fixed effects. Standard errors are clustered at the individual level. * $p < .10$, ** $p < .05$, *** $p < .01$.